# Quantifying the effect of individual heterogeneity in epidemic spread on a network

**Stephen Merry**

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

**Master of Science in Mathematics**

School of Mathematics and Statistics

University of Canterbury

New Zealand

2021

# Contents

# Abstract

The common brushtail possum is an introduced species that, over the last 150 years, has caused significant damage to New Zealand's native flora and fauna. Moreover, brushtail possums are one of the major wildlife reservoirs of bovine tuberculosis, and are thought to be one of the main vectors of the disease to cattle.

In this thesis, we analyse data, collected by Dr Dan Tompkins and his team at Manaaki Whenua - Landcare Research, detailing the interactions between 4 groups of approximately 40 possums over a 6 month period. Further, a strain of Bovine TB was deliberately introduced into the population, and the occurrence of secondary infections were observed.

We construct an individual-based model that models the movements of and interactions between a group of individuals, and produce probabilistic results about the frequency and duration of their interactions. From this individual-based model, we construct networks representing these possum populations, weighted with the frequency of contact between individuals.

After presenting a stochastic transmission model for the simulation of an epidemic on a weighted network, we use this transmission model, coupled with the weighted contact networks mentioned above, to simulate bovine tuberculosis epidemics spreading through a possum population.

# Acknowledgements

First, and foremost, I would like to thank my supervisors, Jeanette McLeod, Alex James and Mike Plank, for their encouragement and guidance throughout my time under their tutelage. Those thanks extend to the entirety of the people who have helped me during my time at the University of Canterbury, including the many wonderful people in the School of Mathematics and Statistics. I would also like to thank Daniel Tompkins and Carlos Rouco for the use of the data they collected.

My postgraduate studies were supported financially by Te Pūnaha Matatini, but their support extends far beyond the monetary. The community that they have built and invited me into with open arms allowed me to enjoy experiences, both professional and personal, that I otherwise would not have done.

Without the support of my parents, James and Ethel, I would not have been able to get where I am today, nor would I have finished this thesis at all. Even though I am 11,803 miles away from you both, I feel like you have never left my side. Finally, I would like to thank all of my friends and family, both here and abroad, for keeping me sane and not letting me give up.

*"A drunk man will eventually find his way home, but a drunk bird may get lost forever."*

Shizuo Kakutani

# Chapter 1

# Introduction

New Zealand has remained geographically isolated throughout most of its history [Cooper and Millener, 1993] and, until the introduction of humans, the only land mammals present were three species of bat [King, 1990]. Humans arrived in New Zealand in two distinct waves; the Māori around 1250-1300CE [Lowe, 2008] and Europeans in 1769 [Liu et al., 1999]. The introduction of humanity to New Zealand, in particular the Europeans, resulted in a parallel introduction of invasive species. The Europeans, accidentally or deliberately, have been responsible for the introduction and establishment of hundreds of invasive species, including land mammals [Veitch and Clout, 2001]. This introduction of invasive animal species has had a devastating effect on New Zealand's native ecosystems. The invasive species not only compete with native species for food and space, but many are also predators of them [McLennan et al., 1996, Innes et al., 2010].

The research in this thesis centres around a specific invasive pest, the common brushtail possum (*Trichosurus vulpecula*), a marsupial of Australian origin that numbered approximately 30 million in New Zealand in 2009 [Warburton et al., 2009]. Aside from the possums' negative effects on native fauna and flora, they impact sectors of the New Zealand economy including the dairy industry, by spreading bovine tuberculosis to cattle [Nugent et al., 2015]. The motivation of the research undertaken in this thesis is to aid ongoing efforts by various organisations to reduce the negative effects of possums in New Zealand. These efforts include programs to reduce or remove possums from an area, or to eradicate bovine tuberculosis from their population. Models are presented that inform on not only the movements of and contacts between possums, but also the intraspecies spread of disease.

## 1.1 Objectives

In this thesis we develop a mathematical model to accurately predict the social structure and contact distribution of common brushtail possum populations. We use an individual-based model, where individuals in our population move around in their environment on a lattice structure, and use the model to produce not only results about the distribution of contacts between individuals, but also a weighted contact network representing the population. We use contact data generated from a real-world possum population [Rouco et al., 2018] to parameterise our model, and compare binary and weighted network characteristics, between networks produced from the model and produced from the contact data.

An epidemic model that simulates the spread of a bovine tuberculosis infection through a possum population is then developed. Using a Gillespie algorithm with parameters obtained from literature, we simulate epidemics on the aforementioned weighted contact networks and compare the model with data collected on real-world bovine tuberculosis infections in the same possum populations as our individual-based model.

## 1.2 Literature Review

This thesis brings together ideas and techniques from both ecology and mathematics. As such, it is necessary to give an overview of the relevant parts of both areas.

### 1.2.1 Ecological Background

Invasion ecology uses a broad range of terminology to refer to species that are introduced to an environment (be it a country, area or ecosystem) that they did not originate from [Colautti and MacIsaac, 2004, Valéry et al., 2008]. We refer to these species as *non-native*. For millennia, humans have been responsible for many of the introductions and establishments of non-native species to new ecosystems [Mack et al., 2000]. However, their introduction can be natural and/or human-driven, illustrated in the example of the Douglas fir tree (*Pseudotsuga menziesii* var. *glauca*), which due to the interaction of factors including human disturbance and climatic variation have been encroaching on the grasslands adjacent to their originating area [Bai et al., 2004].

Invasive alien species are defined as the subset of the non-native species that have a negative effect on the ecosystems, habitats, and native populations of the environments that they invade [Pejchar and Mooney, 2009]. The severity of these negative effects vary, and in addition to negative ecological and economic effects they bring, invasive species can be responsible for the spread of disease [Conn, 2014]. There are cases of invasive species who also provide neutral or even positive effects, but overall the invasion of non-native species has been ranked as the second greatest cause of species endangerment and extinction, behind habitat destruction [Levine and D'Antonio, 2003, Pejchar and Mooney, 2009].

Preventative measures are currently, and have been in the past, used to stop the establishment of new invasive species. For those invasive species who are already in a location, control measures have been used and are in use, in an effort to remove or reduce their populations. However, these measures are expensive, potentially damaging to the native ecosystem, and may unfortunately be unsuccessful. We discuss this in greater detail in the section titled "The effects of invasive species on native ecosystems", and in further detail focusing on common brushtail possums in New Zealand in the section titled "Control measures of possums in New Zealand".

**Introduction of invasive species to native ecosystems**

There are a variety of different routes through which invasive species are introduced to new environments. Many of these routes are human-driven and can be a result of aspects of global infrastructure including the international transport of goods, migration of humans between distant locations, in addition to agriculture or even leisure activities [Perrings et al., 2005, Meyerson and Mooney, 2007, Anderson et al., 2014].

For example, as a centre of international trade and destination for travel over the past centuries, Europe has seen a variety of alien species become established and later invasive, alongside their corresponding negative effects [Keller et al., 2011]. Over 10,000 alien species now established in Europe [Vilà et al., 2010]. Similarly, for the 45 years up to 2005, commercial shipping has been implicated as the source of 60% of new introductions of invasive species to the Great Lakes area of the United States [Horan and Lupi, 2005]. An infamous example of introduction through trade is the zebra mussel (*Dreissena polymorpha*), which has been introduced from its native habitat in the Caspian and Black Sea regions to new environments through avenues including sea trade [Minchin et al., 2002]. The mussel arrived to North America in the ballast water of ships, and to European countries including Ireland by travelling on the hulls of boats [Vitousek et al., 1996, Anderson et al., 2014]. The negative effects zebra mussels have on the countries they invade will be discussed in the section following this one.

Species have been introduced as crops or livestock for the purposes of agriculture [Pimentel, 2011], or to regions whose native flora and fauna would not be sufficient to sustain human populations [Pimentel, 2002]. The golden apple snail (*Pomacea canaliculata*) is a good illustrative example. Native to South America, this mollusk was initially smuggled in Taiwan illegally in 1979/1980 [Naylor, 1996]. From there, it was then introduced to other parts of Asia, including Japan in 1981, and the Philippines in 1982 [Naylor, 1996]. The purpose of its introduction was cultivation as a high-protein food source for consumption domestically, and export internationally [Halwart, 1994, Naylor, 1996]. However, whether accidentally or deliberately, the snail became introduced to the native ecosystem, becoming an invasive pest [Halwart, 1994, Naylor, 1996]. In addition to agricultural purposed, species have been introduced to regions, southern Africa for example, for industries including timber [Le Maitre et al., 2011].

Leisure activities are responsible for 7% of freshwater introductions of invasive species, according to data from

the pan-European project DAISIE (Delivering Alien Species Inventories for Europe) [Gallardo and Aldridge, 2013]. Anderson et al. (2014) gathered data, using questionnaire surveys, about cleaning regimes employed on the equipment of anglers and canoeists in the United Kingdom. They found that not only did a significant minority not clean their equipment between uses, but that this sometimes occurred when returning from overseas. They highlighted in particular the threat of the salmon louse (*Gyrodactylus salaris*), which is a bio-security threat to the United Kingdom, and could possibly be transported from Norway [Anderson et al., 2014].

**The effects of invasive species on native ecosystems**

Invasive species by their nature have a negative effect on the ecosystems they invade [Pejchar and Mooney, 2009]. There are many ways in which invasive species negatively impress themselves upon an environment; a common one is predation of and/or competition with native species. In the Unites States, 400 out of the 958 species listed by the ESA as endangered or threatened are there as a result of competition or predation by non-native species [Pimentel, 2011]. In New Zealand, there are at least 12 species of well-established invasive predatory mammals, many of whom were first introduced during European colonisation [McLennan, 2006]. Stoats (*Mustela erminea*), alongside the weasel and ferret were introduced to New Zealand towards the end of the 1800s in order to control rabbit populations, and now rank amongst these aforementioned invasive predatory mammals [King and Moody, 1982]. They have had a devastating effect on native New Zealand fauna, birds in particular, being implicated in the reduction of Kaka (*Nestor meridionalis*) populations in New Zealand [Wilson et al., 1998].

Ecological effects can also be observed after the introduction of invasive species. Zebra mussels for example, whose introduction to North America and Western Europe was discussed in the previous section, have a negative ecological impact on the countries the invade. For example, their redistribution of contaminants in the water, including cadmium and organochlorine, have resulted in teratogenic effects in the eggs of the tufted duck (*Aythya fuligula*) [Ludyanskiy et al., 1993]. The Seneca River in New York, United States has seen a dramatic water quality changes since the introduction of the Zebra mussel [Effler et al., 1996]. These changes have included a reduction in both dissolved oxygen levels and phytoplankton biomass, in addition to a decrease in pH [Effler et al., 1996].

The introduction of an invasive species can also result in the spread of disease [Mooney et al., 2005]. A

particular disease could be one already present in the environment, but that finds a new host in the invasive species, or could be something brought to the new environment along with the invasive species [Searle et al., 2016]. Mosquitoes are a prime example of this phenomenon. The invasive tiger mosquito *Aedes albopictus* has been shown to be the main vector of Chikungunya virus (CHIKV) during Gabon's 2007 outbreak [Paupy et al., 2012]. Crayfish plague, caused by *Aphanomyces astaci* Schikora, affects crayfish populations globally, though American crayfish are less susceptible. The spread of this plague from the United States to Europe in the 1800s has been largely human-driven [Taugbøl et al., 1993].

Bovine tuberculosis, caused by the bacterium *Mycobacterium bovis* [Morris and Pfeiffer, 1995], is a long standing epidemic in New Zealand, having been recognised as a problem for over 120 years [de Lisle, 1993]. The complex system of interacting species that perpetuates the disease involves cattle, as well as wild populations of deer, possums, ferrets and pigs [Nugent, 2011]. Due to the practice of removing infected herds once infection is detected, thereby stopping transmission and spillover into other species, TB in cattle is typically caused by possums [Nugent, 2011], though ferrets may also hold some responsibility [Ryan et al., 2006]. Possums themselves cause infection spillover into deer, ferrets and pigs. However, due to their large home-range sizes, comparative longevity, and ability to survive years whilst infected, deer have the potential to cause infection to spill back into possum populations [Nugent, 2011].

Invasive species can have a negative economic impact of the environments they invade. This economic impact can manifest in the form of effects on infrastructure. Zebra mussels, for example, have a destructive effect on water treatment and hydro-electric facilities in the Unites States, where they foul intake pipes amongst other equipment [Connelly et al., 2007]. Species invasions may also result in the reduction in usability of economically productive areas. In Florida, United States, hydrilla (*Hydrilla verticillata*) infestations prevent the use of lakes for recreational use [Pimentel et al., 2005]. Yellow star-thistle (*Centaurea solsitialis* L.), which has invaded the rangelands of Idaho, United States, reduces these rangeland's usability for animal grazing [Julia et al., 2007].

Invasive species do not have solely negative effects. Investigation has been conducted into the possibility of using invasive seaweeds, that are otherwise difficult to control or eradicate, as potential pharmaceutical sources [Milledge et al., 2016], or as greener anti-fouling agents for aquatic ecosystems [Pinteus et al., 2020]. Furthermore, non-native species considered harmful to particular organisms are advantageous to other

organisms. For example, many form the majority of agricultural production in some countries. In the United States, 98% of the food system is produced by non-native species, both plant and animal [Pimentel, 2011]. In New Zealand, whilst the European legume shrub Scotch broom has many negative effects, including preventing new forest growth by smothering tree plantings, it has been shown to be beneficial to bee populations [Jarvis et al., 2006].

Control measures have been used to control populations of invasive pests. An infamous example of the use of chemical for invasive population control is the use of Dichlorodiphenyltrichloroethane, also known as DDT. Used as an insecticide, its devastating effects on ecosystems, including population declines in species such as the bald eagle (*Haliaeetus leucocephalus*) [Sharpe and Garcelon, 2005], weren't realised until years after its first use [Jarman and Ballschmiter, 2012].

Biocontrol agents are also a common control measure used. In 1993, Jolly defined a biocontrol agent as "the use of one organism to reduce the population of another" [Jolly, 1993]. The use of biocontrol agents is not novel, in New Zealand particularly, and has had devastating, unintended results. Stoats (*Mustela erminea*), for example, were introduced to New Zealand for the purposes of controlling the rabbit population, but have since become a significant invasive pest [Basse et al., 1999, Veale et al., 2012]. There are success stories however, with the successful control of the mist flower in New Zealand through the introduction of the white smut fungus and the gall fly [Barton et al., 2007].

Pathogens are also a viable source of biocontrol agents. A strain of rabbit haemorrhagic disease virus (RHDV1), for example, was illegally introduced into New Zealand in 1997 by farmers, for the purpose of controlling the wild rabbit population [Parkes et al., 2002, O'Hara, 2006]. RHDV1 has been an effective population control for rabbits, and has been bolstered by the nationwide release of a Korean strain of the virus (RHDV1 K5) was undertaken by government and local councils at 150 sites across the country in 2018 [Manaaki Whenua - Landcare Research, 2016, Manaaki Whenua - Landcare Research, 2018, Howe, 2018]. Fertility manipulation is another route. Infertile screwworms, made infertile through irradiation, were released over the island of Curaçao, leading the insect's decline [Carey, 2018].

**The common brushtail possum in New Zealand**

The common brushtail possum is an arboreal marsupial native to northern, eastern, south-western, and central Australia [Menkhorst et al., 2004]. A typical member of this species weighs 2–4 kilograms, and is of a similar size to a cat [Clout and Ericksen, 2000].

Between 1837 and 1924, the importation of common brushtail possums to New Zealand from Australia by European settlers took place on at least 35 separate occasions [Pracy, 1974, Clout and Ericksen, 2000], with the majority occurring in the 1890s [Pracy, 1974]. The purpose of these importations was to farm the possums for their fur [Batcheler and Cowan, 1988, Clout and Ericksen, 2000].

Liberation events, where imported or New Zealand-bred possums were set free (or liberated) into the environment, took place at various sites across the country, starting in late 1850s [Pracy, 1974]. The liberation of possums was carried out by government agencies, private individuals and organised groups called Acclimatization Societies, with the goal of establishing a wild possum population for hunting and for profit [Batcheler and Cowan, 1988, Clout and Ericksen, 2000]. Between 1865 and 1926, 127 recorded liberations took place, with 19 of these made by government agencies [Clout and Ericksen, 2000]. The number of individuals involved in each importation or liberation event is unclear or unknown [Pracy, 1974]. This practice was completely prohibited from 1922, but was continued illegally by trappers until the 1980s [Clout and Ericksen, 2000, Department of Conservation, 2004].

In 1947, in response to scientific evidence of the impact of possums on native forests, the government officially recognised them as a major pest, and introduced penalties for the possession or liberation of possums [Clout and Ericksen, 2000]. Possums are now endemic to 95% of New Zealand [Sidhu et al., 2002], with the most recent population estimate (in 2009) providing a figure of approximately 30 million [Warburton et al., 2009].

Possums are nocturnal, solitary animals that spend roughly 1% of their time in contact with other members of their species [Day et al., 2000]. The majority of their day is spent in their dens, usually only leaving their dens around 30 minutes after sunset [Cowan and Clout, 2000]. The exceptions to this are during illness or winter, when a sick or starving possum may emerge in the afternoon in search of food [Cowan and Clout, 2000]. Time outside their dens is spent travelling, grooming, feeding, interacting, or sitting immobile [MacLennan,

1984, Cowan and Clout, 2000]. When two possums do come into contact, their behaviour can be classified into four categories: approach and leave, agonistic (where two possums fight), affiliative (where two possums may form social bonds), and mating [Day et al., 2000].

The negative effect of possums on the ecosystem of New Zealand can be divided into three categories: their effect on native vegetation, their effect on native animals, and the transmission of bovine tuberculosis to cattle [Nugent, 1995, Landcare Research, 2000, Landcare Research, 2002].

Each day the possum population consumes, in total, 21,000 tonnes of foliage [Nugent, 1995]. This consumption causes catastrophic dieback, as well as a change in the composition of forests through the consumption of certain preferred species [Campbell et al., 1990, Nugent, 1995]. They may also have a negative effect by inhibiting regeneration; however, the impact of possums on regeneration is not well understood [Campbell et al., 1990]. Possums also threaten New Zealand's native fauna. They not only compete with other species for vegetation, but also predators of them. They kill or eat the eggs, chicks, and adults of at least six native species, including kea, kōkako and kūkupa [Landcare Research, 2002].

**Bovine tuberculosis in possums**

Bovine tuberculosis (TB) is endemic to the common brushtail possum in New Zealand [Morris and Pfeiffer, 1995]. Possums are highly susceptible to the disease [Ramsey et al., 2009], and are the major wildlife reservoir of TB in New Zealand [Landcare Research, 2000]. Since 1967, when bovine tuberculosis was first discovered in New Zealand possums [Ekdahl et al., 1970], various areas in New Zealand have been designated as Vector Risk Areas, meaning that TB is established in possum populations in that area. These areas constituted 7.51 million hectares (approximately 28% of New Zealand's land area) in 2019 [OSPRI New Zealand (OSPRI), 2019].

Naturally-occurring bovine tuberculosis infection has been studied in New Zealand possums across the country, for example by Brockie et al. in 1987, by Cooke et al. and by Pfeiffer et al., both in 1995. There are several possible routes of TB infection in possums; transmission during mating, fighting, sequential den sharing, and mother to pouch young transmission, however the exact mechanism is not well understood [Jackson et al., 1995, Ramsey et al., 2009, Richardson, 2016]. Clinical signs of infection begin to show 2-3 months post infection [Roberts, 1992, Corner et al., 2002]. Once clinical signs appear, possums typically die within 4-5

months [Ramsey and Cowan, 2003].

The study of TB in possums has been conducted through both the monitoring of populations with already established infections, and through the deliberate infection of individuals. In 1999, Cooke et al. examined the differences between experimental and wild infections. They found that infecting possums through an endo-bronchial route may not be the most appropriate method for studying the epidemiology of TB in possums, as the pathogenesis was markedly different [Cooke et al., 1999]. Corner et al (2003) showed that infecting possums via the conjunctiva produced symptoms more similar to wild infections than the previous intratracheal inoculation. Whitford et al (2014) infected individuals percutaneously, in a previously TB-free population. They demonstrated possibly the first case of TB transmission from experimentally infected possums, suggesting that percutaneous infection improves on previous methods when examining TB transmission in the wild [Whitford et al., 2014].

Possums are thought to be one of the main vectors in New Zealand of TB to cattle [Nugent et al., 2015]. Results from experiments performed by Paterson and Morris in 1995 showed that behaviour similar to that of terminally ill possums attracts the attention of cattle, who approach and investigate the possums by sniffing them, touching them with their noses, and licking them. These actions expose them to possible TB infection [Paterson and Morris, 1995]. Infected cattle have to be slaughtered, hence the transmission of bovine tuberculosis to cattle is a concern to the beef and dairy industries of New Zealand. In turn, TB transmission to cattle can negatively impact New Zealand's economy, as the dairy, meat and wool industries account for a significant percentage of New Zealand's export (approximately 61.5% of all exports in 2019 [Ministry for Primary Industries, 2019]).

### Control measures of possums in New Zealand

In an effort to mitigate the negative effects of the common brushtail possum's presence in New Zealand, various methods, including hunting and poisoning, have been used to reduce or eradicate the possum population, or eradicate TB from the possum population. For example, a bounty system was run by the government between 1951 and 1961 that resulted in 1 million possums killed per annum in the late 1950s [Clout and Ericksen, 2000].

To eradicate TB from New Zealand, the National Pest Management Strategy for Bovine Tuberculosis was announced by the government in 1998 [Michael Hardie Boys, Governor-General, 2016]. Now named the National Bovine Tuberculosis Pest Management Plan, the current strategy aims to eradicate TB from New Zealand by 2055, having eradicated it from cattle and deer by 2026, and from possums by 2040 [Michael Hardie Boys, Governor-General, 2005].

The agency managing this plan of nationwide eradication, TBfree New Zealand, is using various methods of pest management to reduce or eradicate the possum population. By reducing the possum population by approximately 70% and maintaining that level, TB in possums is predicted to die out within 5-10 years [Landcare Research, 2000]. These methods include live capture and kill traps, as well as poisons such as 1080 and cyanide [Warburton and Livingstone, 2015, Byrom et al., 2016]. These methods are time and cost intensive, requiring continual renewal and upkeep of resources [Jolly, 1993, Morgan and Hickling, 2000]. Additionally there is public opposition to the use of the poison 1080 for the control of pest populations, due to concerns over factors including human safety and collateral damage to native species [Green and Rohan, 2012, Morriss et al., 2020].

Alternative, but not necessarily replacement, biological forms of possum population management are in development or consideration. These include the use of vaccinations to disrupt fertility [Cowan, 2000] or to provide immunity to TB. In 2009, Tompkins et al. showed that the use of the *M. bovis* Bacille Calmette-Guérin (BCG) vaccine on possums had a significant effect on the incidence of TB in free-living possums. Further experiments conducted in 2013 showed that protection of possums against TB can be provided through a single dose of a BCG vaccine [Tompkins et al., 2013]. These methods would need less continual input or intervention, but would require monitoring of their effectiveness [Jolly, 1993], and can be cost-effective, overall.

The research carried out in this thesis will aid in the development of control methods such as vaccination and the use of biocontrol agents. By building models of the mechanisms of contact and disease spread in a possum population, insight may be gained that could allow for the targeting of specific individuals for vaccination or removal, in order to achieve the goals of nationwide TB eradication or possum population control more efficiently. Furthermore, with appropriate adjustment of the models, they could be used for any invasive species, which would be of much use to the Department of Conservation and its goal of a predator free New Zealand in 2050 [Department of Conservation, 2020].

### 1.2.2 Mathematical Background and Motivation

In this thesis, a variety of mathematical techniques are used. This section provides a brief overview of these techniques, reviews literature that uses these techniques to approach problems similar to those we investigate.

#### Mathematical modelling of biological processes

A plethora of models have been developed to model a large range of biological processes (see [Gail and Boone, 1970, Lovely and Dahlquist, 1975, Johnsson et al., 1996, Hill and Häder, 1997, Haydon et al., 2008, Karaman et al., 2016] and references therein). These models tackle activities occurring at different spatial scales, over a variety of temporal scales.

An individual based model (IBM) is a particular class of model in which, rather than modelling a population as a whole, be it of microscopic or macroscopic organisms, each individual is modelled separately. Of particular relevance to this thesis is a specific family of IBMs, namely random walks. A *random walk* uses a sequence of random vectors, drawn from a given distribution, to describe the movements of an individual through an area over a specified time period.

Many different biological processes have been modelled using a random-walk framework. Karaman et al. (2016), for instance, showed that a continuous-time random walk can allow for better differentiation between severe and less-severe paediatric brain tumours, which in turn aids in non-invasive diagnosis. Johnsson et al. (1996) investigated the effects of micro-gravity on the growth of garden cress roots, finding that the initial stages of growth follow aspects of a random walk. Gail and Boone (1970) investigated the movements of mouse fibroblasts. By comparing the fibroblast trajectories against random walk models they determined that, given a large time-step, a random walk model was appropriate for modelling the movements of these fibroblasts.

This is by no means the only instance of a random walk model being applied to cellular movement. Hill and Häder (1997), for example, modelled tracks of algal cells, using a continuous-time (that is, the time steps are non-discrete), correlated random walk. In a correlated random walk, the movements taken are directed towards a particular location, for example a food source [Kareiva and Shigesada, 1983]. Lovely and Dahlquist (1975) produced a three-dimensional random walk model for the movement of bacteria, with a view

to comparing with data collected the preceding year on *Escherichia coli*.

The individual movements of animal species have also been modelled using a variety of random walk models. For example, Haydon et al. (2008) fitted a random walk model to data collected on the behaviour of elk (*Cervus canadensis*) who were trans-located to an area of Ontario, Canada, and tracked over a 4-year period. From the data collected, the elks' movements could be assigned to two different behavioural states, one when they were occupying one particular area, or 'camp', and the other when they were ranging far from these camps [Haydon et al., 2008]. Each behavioural state was modelled by a different random walk.

Ríos-Uzeda et al. (2019) analysed three marsupial species in Brazil, finding that a specific type of random walk, known as a Lévy flight or Lévy walk, was best at approximating their small-scale movements when introduced to an unfamiliar environment. The application of random walk models extends beyond mammalian species. Kareiva and Shigesada (1983) model the movements of butterflies (*Pieris rapae* and *Battus philenor*), using correlated random walks. Findings from Sims et al. (2014), who analysed the tracks left behind by Eocene-era worm-like animals, suggested that the movements of these foraging animals can be modelled using Lévy walk-type models.

A major difference between the work conducted in this thesis and many of the studies that use random walk models is the the form of the data used. Typically, the data collected has an explicit spatial component (see [Kareiva and Shigesada, 1983, Haydon et al., 2008, Postlethwaite and Dennis, 2013, Reynolds et al., 2017] and references therein). In Reynolds et al. (2017) for example, limpets *Patella vulgata* and *Patella rustica* were tracked, forming a data-set of their explicit movements. In contrast, the data examined in this thesis consists of contact records, and has no spatial component.

In this thesis, possum movements and contacts are modelled using an unbiased random walk on a square lattice structure. Here, possums move one point on the lattice at a time, with no preference towards a particular location. A *contact* occurs when two individuals occupy the same point on the lattice.

Modelling possum movement and contact using a random walk has precedence. Postlethwaite and Dennis (2013) used a multiple-random walk. Here, possums' movements can belong to one of three different behavioural states, or types of behaviour. Each is expressed as a different random walk. However, their work differs from

ours in three key areas. Firstly, individuals in our model remain in the same behavioural state throughout, following the same random walk at all times. Secondly, the data used to verify their model is fundamentally different to the data used in this thesis. Whereas our data consists of contact records between possums, with no spatial information, the data set in Postlethwaite and Dennis consists solely of GPS tracking data, with no contacts being recorded. This highlights the final difference, that their model, whilst producing results about the movements of possums, outputs no information about the contacts between them.

James et al. (2017) provides a model that combines both the movements and contacts between possums in the same populations that we study in this thesis. As with our model, it is assumes that individuals have a fixed area in which they live, feed, and breed, known as the home-range, and that the centres of these home-ranges are distributed according to some point process. The movements of possums are described using a mean-reversion random walk, where possums' movements tend towards the centre of their home-range. As the distance from the centre increases, movements towards the home-range centre are more likely. By defining a contact as two possums being within a threshold distance of one another, their model produces measures of the time two individuals spend in contact, given the distance between home-range centres. Mathematical objects called binary contact networks are produced to represent the possum populations, this will be expanded upon in the next section.

To summarise, the model we produce will model the movements of possums, with no preference towards a particular location or direction, unlike those of James et al. (2017) and Kareiva and Shigesada (1983), and will follow the same random walk throughout. Like James et al. (2017), we produce measures that describe the contacts between possums, but our model will focus on the likelihood of a contact occurring, rather than the time spent in contact.

### Network representations of social structure in a population

The underlying social structure of a population (assumed to be of size $N$) is an important tool when analysing contacts or predicting disease spread, and can be represented using a network. A network consists of two sets: one of points or vertices (also called nodes), and the other of links or edges between vertices. Although in a real-world context a network can represent any number of different entities and their respective relationships, in this thesis, the networks produced and used have vertices representing individual possums, and the edges indicating that a contact has occurred between two individuals.

Networks, created using empirical data sets of contacts between individuals, have been produced for populations including bumble bee (*Bombus impatiens*) colonies in the case of Otterstatter and Thomson (2007), and Tasmanian devils (*Sarcophilus harrisii*) in Hamede et al. (2009). Both of these studies produced networks using explicit contact data gained from video-tracking and proximity collars respectively.

There are examples of empirical contact networks being produced from data on the explicit contacts between possums (see [Corner et al., 2003b, Ji et al., 2005, Porphyre et al., 2008, James et al., 2017, Rouco et al., 2018] and references therein). In a similar fashion to the data we use in this thesis, Ji et al. (2005) produced contact networks of wild-living common brushtail possums from data collected by proximity loggers. Edges in their networks exist if a contact was recorded during the study period. Indeed, the empirical contact data we use has been used previously to produce contact networks. James et al. (2017) and Rouco et al. (2018) produce contact networks from the same data set, with edges existing between pairs if there was at least one contact between them during the study.

Although empirical contact networks of common brushtail possums have been produced before, not all have been produced through the use of explicit contact data. In Corner et al. (2003), an edge exists between a pair of possums if they share the same den at any point during the diurnal hours of the study. In addition to the edge condition, Corner et al. (2003) also differ from the data set we discuss in Chapter 2, in the respect that the possums studied were captive, meaning that their social structure is not necessarily the same as the free-living possums we study here. Porphyre et al. (2008), on the other hand, produce networks of free-living possums. Edges in the networks of Porphyre et al. (2008) exist pairs if either sequential trap use or trapping in the activity range of another possum was documented, whereas edges in our empirical contact networks are contingent on explicit contact.

Contact networks can also be produced from models. In these models, vertices represent individuals or entities, and a condition for the existence of an edge is made explicit. For example, Wang and González (2009) produce network models of human populations, where edges are determined by individual's proximity to cell towers. This approach differs to the networks from our models as, though edge existence is determined by proximity, the proximity is to an external object, rather than to other individuals. Rolls et al. (2015) produce network models of the contact structure of an American high school over a single day. Edges exist between vertices

representing students, staff and other individuals if they were in close proximity to one another.

Network models of possums populations have also been produced. James et al. (2017) used results from their mean-reversion random walk model (described in a previous section) to produce a contact network that model the contact structure of the same possum population we study. However, whilst edges in our network models exist contingent on a contact occurring, edges in the networks of James et al. (2017) exist if the estimated total contact time exceeds a given threshold value.

However, by representing the social structure between individuals using only a binary network, the heterogeneity of their contacts is not modelled. A solution is to assign to each edge a number, referred to as the edge weight, forming a *weighted network*. These edge weights typically represent the strength of the contact. Many different scenarios and contexts have been modelled using weighted networks. In Newman (2001), the heterogeneity of collaboration between academics is represented by a weighted network, with edges weighted with the number of papers two academics have collaborated on. Networks representing activities including transport have also been modelled. Both Bagler (2008) and Barrat et al. (2004) model air traffic networks, creating networks air transport data, weighting edges with measures indicating the level of traffic handled by airports.

Of relevance to the work done in this thesis is the use of weighted networks to represent the strength of contacts between individuals in a population. Drewe et al. (2011) produce empirical weighted contact networks from contact data. Edges are weighted by the number of contact events of a certain behaviour (grooming, for example). Onnela et al. (2007) produce networks from mobile phone data, weighting edges in two ways: by the total duration of calls, and total number of calls.

### Compartmental Epidemic Models

First described in 1927 by Kermack and McKendrick, in compartmental epidemic models a population is divided into disjoint groups, called compartments, according to their current epidemic status. Each stage of the epidemic is represented by a single compartment, and members of a single compartment are assumed to be at the same stage of the epidemic. Each individual is in one and only one compartment at any given time, but can transition from one to another [Blackwood and Childs, 2018].

The most basic compartmental model has only two states; Susceptible, where individuals are not currently

infected, and Infected. Susceptible individuals move to the Infected state at a given rate. As discussed in McCallum et al. (2001), a variety of functions are used to describe this rate, typically involving a combination of the numbers of Susceptible and Infected individuals. This model is referred to as an SI-model, and has been used to model diseases including bovine tuberculosis in common brushtail possums [Roberts, 1996]. Diseases including influenza have been modelled using an SIS-model, where individuals return to the Susceptible state after recovering from a disease (see [Abouelkheir et al., 2017], for example). Introducing more states into the model adds greater complexity. An SEI-model incorporates an Exposed compartment, where individuals are infected but not yet infectious, and therefore unable to infect others. Models including the SIR-model and SEIR-model build on the SI-model and SEI-model respectively by introducing the removal or recovery of individuals, thereby creating a Removed state. Many more variations and expansions also exist (see [Bichara et al., 2014, Thompson, 2016, Ehrhardt et al., 2019, Hackenberger, 2019] and references therein for examples).

Compartmental models have been used to model the spread of numerous diseases, infections, and epidemics; across different species, different spatial scales, and over different time periods. We discuss the use of compartmental models in this different scenarios in the paragraphs subsequent to this one.

This class of model has been used to model many of the major diseases that affect humanity (see [Ferguson et al., 2003, Lekone and Finkenstädt, 2006, Hailegiorgis and Crooks, 2012, Blyuss, 2016, Getz et al., 2019, Ameri and Cooper, 2019] and references therein). Various outbreaks of Ebola hemorrhagic fever in West Africa have been modelled using SIR-models or their extensions. For example, Lekone and Finkenstädt (2006) developed an SEIR-model for the 1995 outbreak in the Democratic Republic of the Congo, using a time-dependent transmission rate which is initially constant, but decays exponentially after control measures are introduced. The 2014–2015 outbreak of Ebola has also been modelled using compartmental models of various flavours [Yang et al., 2015, Kucharski et al., 2015, Getz et al., 2019]. Kibona and Yang (2017) modelled the spread, starting in 2016, of the Zika virus in Brazil.

Many bacterial diseases have also been modelled. Ameri and Cooper (2019), for example provide both a classical and a network-based SEIR-model of whooping cough, caused by the bacterium *Bordetella pertussis*, in Nebraska, United States. They used data consisting of weekly reports of whooping cough cases from 2000 to 2017, showing that the classical SEIR-model overestimated the number of people infected, whereas the network model provided a closer estimate. Hailegiorgis and Crooks (2012) produced an SEIR-model, with

additional external infection sources, for the spread of cholera in East African refugee camps. Blyuss (2016) modelled meningococcal meningitis in Africa, with a carrier state of individual in lieu of exposed. Here, carriers are able to transition to the removed state without first becoming infectious.

Compartmental epidemic models have even been used to inform about potential outbreaks. Ferguson et al. (2003) investigated various compartmental models for smallpox spread following an bio-terrorist attack, as well as the most appropriate control method to implement. Their conclusion was that no one model or control method could be deemed the most appropriate. Instead, a variety of different compartmental epidemic models, with different epidemic stages and control methods, should be used to prepare for a smallpox outbreak.

Epidemics in animal and plant species are also often studied using compartmental models (see [Roberts, 1996, Gilligan et al., 1997, Barlow, 2000, Keeling et al., 2001, Ferguson et al., 2001, Nielsen et al., 2007, Reynolds et al., 2014] and references therein). Challenges not as apparent with human diseases appear when modelling animal or plant populations, including contact tracing, the wider social contact network, and the estimation of certain epidemic parameter values. Nielsen et al. (2007) estimated parameters for a *Salmonella* Dublin infection in dairy calves, through the use of an SIR-model and real-world data about the calves' infection statuses. The spread of Foot and Mouth disease in the United Kingdom has been modelled using compartmental models in both Keeling et al. (2001) and Ferguson et al. (2001). Reynolds et al. (2014) used an SEIR-model, informed by empirical data, for the spread of influenza in swine herds, where each weekly stage in the swine production process is represented by a separate set of SEIR compartments. They found that influenza was maintained in the population through infectious piglets. Gilligan et al. (1997) produced an SIR-model of the stem canker disease that affects potatoes, which included terms describing the effect of infected stems on the production of new, susceptible stems, and of the increased resistance of stems to the disease.

There is precedent for using compartmental models to describe the spread of TB in a possum population. For example, Barlow (1991a) describes an SEI-model, with birth and death rates, as well as pseudovertical transmission from mother to offspring. In this model, the death rate is a combination of both density-dependent (that is dependent on the population density) and density-independent terms. Roberts (1992) provides a similar model, where instead the death rate is solely density-dependent. Roberts (1996) describes an SI-model, where infection can occur through direct contact, indirect contamination or pseudovertical transmission. Two types of population control, culling and vaccination, are also modelled. Subject to choice of parameter

values, Roberts shows that culling the possum population to a low density is the most efficient method of TB eradication.

These models are well-established, and some have been used to determine the appropriate course of action for eliminating TB in New Zealand possums [Landcare Research, 2000]. However, they differ from the approach we are taking in three key areas.

The first is the manner in which infection is assumed to spread. In Barlow (1991a), Roberts (1992), Roberts (1996) and Barlow (2000), the rate at which susceptible individuals become infected is expressed in terms of the product of the proportions of susceptible and infectious individuals. There is some variation. In Roberts (1996), this rate also incorporates a density-dependent contact rate function. All of these models assume population mixing is occurring, where each individual has the possibility of coming into contact with any other. However, when looking at the contact data we have (see Chapter 2), it becomes clear that, in our case at least, this assumption does not hold. There are more pairs of possums who never meet than those who do.

Therefore, in order to better model the situation, we need to incorporate an explicit contact structure. A similar approach is given in Ramsey and Efford (2010). Here each possum is represented by a point in space and the rate at which bovine tuberculosis infection is transmitted is given as a function of the Euclidean distances between these points. As such, infection is unlikely to be transmitted between individuals living far from each other.

The second key area in which our model differs to existing models is that of deterministic versus stochastic approaches. Barlow (1991a), Roberts (1992), Roberts (1996) and Barlow (2000) all provide deterministic models of TB spread through possum populations. Though Barlow (1993), providing an expansion on the model of Barlow (1991a), incorporates a stochastic element of TB spread through migration, the majority of the model is still deterministic. We instead employ a stochastic approach, as deterministic models are not always appropriate for small communities like the populations we model [Britton, 2010].

The final key area is population dynamics. In most models, there is an assumption of birth, non-disease related death, migration, and pseudovertical transmission. However, we do not incorporate these into the model we produce. This is due to the comparatively short timescale we model, where these factors have little

bearing on the results.

In this thesis, we employ an SEIR-model. This type of model was chosen because TB does have a latent period, requiring the inclusion of an exposed compartment, and it is lethal to possums, thereby requiring us to model the removal of individuals. As mentioned previously, we assume that the population is static, that is, there are no births, migration, or non-disease related mortality. Further, we assume that the only infection transmission route is through direct contact, no pseudovertical transmission nor indirect contamination. Finally, we use the explicit contact structure of the population in order to model the epidemic. This in contrast to the previous models, where although there was some contact heterogeneity modelled, it was not made explicit, nor was it treated on an individual-level.

**Epidemic models on networks**

Many epidemic models, especially those concerned with large populations, use compartmental models with a well-mixed assumption (see for example [Lekone and Finkenstädt, 2006, Ameri and Cooper, 2019] and references therein). This assumption posits that each infectious individual has a chance to infect every susceptible member of a population. However, this assumption often is not sufficient to capture the heterogeneity of the contacts. Nor is it always realistic, as particular individuals in a population may not come into contact at any point, preventing infection being directly transmitted between them. A solution to the shortcomings of the well-mixed model is to use a network framework.

Modelling disease spread using a contact network, describing a population's social structure, is a common method (see Newman (2002) and Keeling and Eames (2005) for an overview). Incorporating the contact network of a population allows for an epidemic model to utilise the explicit contact structure to better model the spread of a disease. Many diseases of humans lend themselves to this approach. Examples include Robineau et al. (2017). In this paper, an SI-model of the transmission of human immunodeficiency virus (HIV) is simulated on theoretical networks representing sexual contacts between men living in Paris, France, in order to examine the effect of pre-exposure prophylaxis on the spread of HIV.

González-Parra et al. (2015) use random networks to model the spread of the influenza A virus (AH1N1) virus in the Venezuelan state of Nueva Esparta. By fitting their SEIR-model to data collected on AH1N1 infections, they showed that, in comparison to a classical well-mixed compartmental model, theirs better

replicates aspects of the data, including the irregularity of the data and the numbers of people infected [González-Parra et al., 2015]. Similarly, Acedo et al. (2011) use random networks, alongside a compartmental SIRS-model, where individuals can become susceptible even after recovering from a disease, to model the spread of respiratory syncytial virus (RSV). They fit their model to hospital data collected on infections in the Spanish region of Valencia during the years 2001 to 2004. Networks produced from empirical data have also been used. In the cases of both Stehlé et al. (2011) and Takaguchi et al. (2013), interactions between attendees at conferences and a Dublin gallery were collected using proximity detecting badges worn by people. Dynamic (or non-static) empirical networks were then produced from these data sets and epidemic spread simulated on them. In Stehlé et al. (2011), an SEIR-model was used, and in Takaguchi et al. (2013) an SI-model was used.

Animal diseases have also been studied using a network lens (see [Craft et al., 2009, Drewe et al., 2011, Laager et al., 2018, Wilson-Aggarwal et al., 2019] and references therein), however problems arise that are not apparent when tackling human diseases, for example data collection. Wilson-Aggarwal et al. (2019) presents a network epidemic model that shares many similarities with ours. They build contact networks of free roaming, but owned, dog populations in Chad. Each of the 108 dogs were fitted with collars that had both GPS and proximity logging capabilities. They simulated an SEIR-model on weighted contact networks, where edges were weighted with the average number of seconds spent in contact. In addition, as a null model, random networks were also used. The purpose of this was to demonstrate the necessity of including contact information when modelling epidemics, especially rabies, which is present in the region of the study. There are differences to the scenario we study and model. They collected no epidemic data and their study period was less than one month.

Craft et al. (2009) modelled the disease dynamics of canine distemper virus (CDV) in Serengeti lions. They built a network model representing lion prides and isolated individuals. They then simulated a CDV epidemic on these networks and compared the results of a "geographically restricted subset...resembling the study population". Drewe et al. (2011) gives an SEIR-model of TB spread in meerkats. Simulations are carried out on empirical weighted contact networks, and numbers of individuals in each compartment at each time interval are compared with the real-world status of the meerkats. However, both of these studies only use empirical contact networks, produced using data.

A common problem faced when studying epidemic spread is the concept of missing data, be it caused by missing individuals, missing links between those individuals [Ghani et al., 1998], or missing infections [Nishiura et al., 2011]. This problem is especially relevant to the study of epidemic spread through a social network, as the omission of data can cause the network structure to be different to that of the underlying network [Kossinets, 2006, Smith and Moody, 2013].

As we will discuss in Chapter 2, there are unknown infection pathways in our empirical possum contact networks, indicating the need to model 'unseen' (that is uncollared) possums. As such, we require a model that produces networks representing theoretical possum populations that are similar to real-world populations with respect to certain metrics, but also model individuals and contacts not recorded in the data. An approach similar to the one we use is also used in Laager et al. (2018). They, like Wilson-Aggarwal et al. (2019), present a model for the spread of rabies through dog populations in Chad. They have a spatially-explicit contact network model that creates a social network that, when a subgraph is compared to the empirical network, has a similar degree distribution. This social network includes 'unseen' dogs, but the method of construction differs from ours. They add edges in their networks by sampling from Poisson distributions. They go on to simulate an SEIR-model on the networks.

## 1.3 Thesis structure

This thesis is split into six chapters. For each, apart from the current chapter, we give an overview of the purpose, content and results.

**Chapter 2: Data on common brushtail possum contacts and bovine tuberculosis infection**

In this chapter, we describe two data sets, produced from a field study that was conducted by researchers at Manaaki Whenua – Landcare Research. Possums in four different populations, living in the Orongorongo Valley in New Zealand's North Island, were trapped. Members of these trapped possums were fitted with proximity loggers. Contact data was then collected on these collared possums. Concurrent to this, members of each population were deliberately infected with bovine tuberculosis. Secondary infections subsequently found were documented, and form the second data set.

We detail first our cleaning process for the data, removing corrupted or invalid records, in addition to daytime records, as our later work produces models only of nocturnal behaviour, and errant or isolated records. From our cleaned first data set, we produce binary networks, with vertices representing possums and edges between pairs in contact, as well as measures of the length and importance of contacts between the possums.

**Chapter 3: Using individual contacts to make weighted networks**

In Chapter 3, we present the creation and implementation of a two-dimensional individual-based model for possum movements and contacts. Each individual follows a random walk, and can come into contact with others. From the model, measures of the length and importance of contacts can be produced, which have analogous values in the empirical data. The contact measures we obtain from the model depend on the distance between the individuals' denning sites, and a parameter that determines the speed of an individual's spread through its environment.

We also detail three methods for producing collared populations of individuals. By using the empirical data, we estimate the value of a parameter for our model, the diffusion coefficient, that then allows us to estimate the home-range sizes of our individuals. Furthermore, we can also produce binary and weighted networks representing model populations, and whose network characteristics are similar to those from the data.

**Chapter 4: A stochastic transmission model for the spread of an epidemic through a network**

In Chapter 4, we present our stochastic transmission model. After detailing the model, we investigate the effect of altering epidemic parameters, network characteristics and initial conditions on the outcome of epidemic simulations.

**Chapter 5: Fitting a stochastic transmission model for the spread of bovine tuberculosis through a possum population to real-world data**

We use the weighted contact networks from Chapter 3, alongside the stochastic transmission model of Chapter 4, to model the spread of bovine tuberculosis through a possum population. We then estimate transmission coefficient, a parameter which relies heavily on the underlying network architecture for its determination. Once we have estimates of our transmission coefficient, we simulate epidemics on our model networks and compare results with the data set of primary and secondary infections, detailed in Chapter 2.

**Chapter 6: Concluding remarks**

In this chapter, we review our work, highlighting the main results and their implications. We also discuss the limitations of the data sets we have, in addition to our models, and detail the scope of work that could be conducted in the future.

# Chapter 2

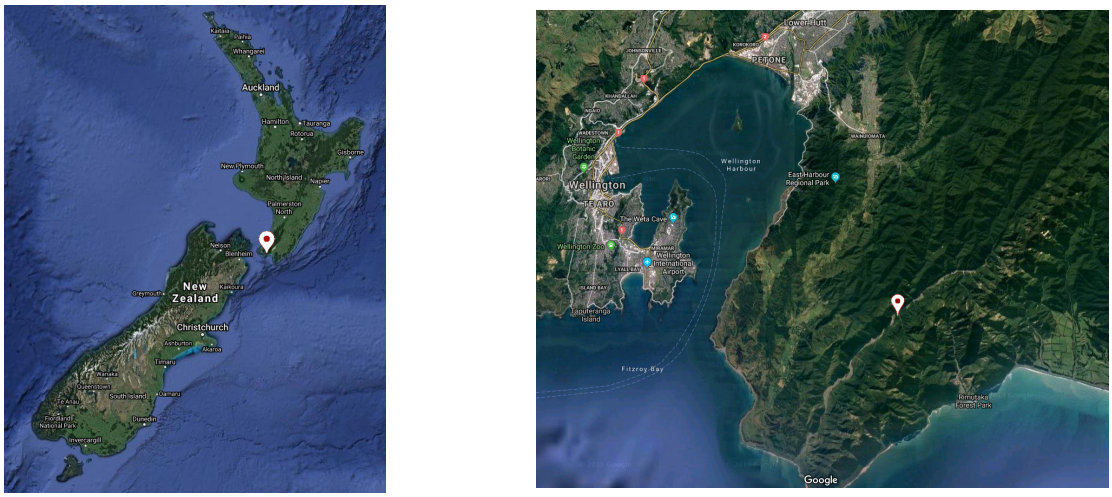
# Data on common brushtail possum contacts and bovine tuberculosis infection

## 2.1 Overview of the field study

The study was conducted over three field seasons (April to December 2012, August 2013 to February 2014, and June 2014 to March 2015) by Dr Daniel Tompkins and his team at Manaaki Whenua – Landcare Research. The aim of the study was to investigate home-range characteristics of common brushtail possums, including the home-range size and overlap between individuals' home-ranges. Individuals in four distinct populations of possums were trapped, a subset of those were fitted with proximity logging radio-collars, which recorded contacts between these individuals. Members of each collared sub-population were deliberately infected with a strain of bovine tuberculosis (TB) that was different to the strains found naturally in the area, and the spread of the strain through the population was monitored, with focus on the influence of factors including population density and sex ratio. Social contact networks were constructed, to provide information about population structure and TB disease spread.

### 2.1.1 Study details

The study population of 313 distinct possums inhabited 1200 hectares of mixed beech-podocarp forest, in the Orongorongo Valley on the North Island of New Zealand (41°21′S 174°58′E, Fig. 2.1). These possums did not comprise the total possum population of the area. There were additional possums were also trapped, but for whom no data was recorded. Additionally, other possums potentially inhabited the area but where never interacted with. The study area was subdivided into four non-contiguous study sites, A, B, C, and D, running north to south. There was a distance of 518 metre between sites A and B, 350 metre between B and C, and 50 metre between C and D. A hillside also separated sites C and D. Each site consisted of a 13 hectare square grid of 100 baited traps, placed at 40 metre intervals. Grieve wire cage traps, of size 60 cm × 26 cm × 28 cm, were used, with spring-assisted folding doors triggered by a pendulum bait hook.




(A) (B)

Figure 2.1: **Location of the study sites.** The white dot indicates the approximate location of the study sites in the Orongorongo Valley in New Zealand's North Island.
Maps of Aotearoa New Zealand. From Google Maps, by Google
(Fig. 2.1A, https://www.google.co.nz/maps/@-41.3129914,174.9049963,52417m/data=!3m1!1e3.
Fig. 2.1B, https://www.google.co.nz/maps/@-41.3789739,173.388883,1675639m/data=!3m1!1e3.)

Trapping sessions, lasting four consecutive nights, were conducted at regular intervals during the seasons. On the first night, each trap was baited and set. Each subsequent night, traps were visited, sprung traps were reset and previously trapped possums released. When first trapped, possums were anaesthetised and tagged in one or both ears. Each ear-tag had a unique ID. Whenever a trap was sprung, the trap location and ear-tag ID(s) of the trapped possum were recorded. During the first trapping sessions of each season, a subset of adult possums on each site were collared with proximity loggers, designed to detect other collars within a 1 metre distance, and record contacts between individuals. If a possum pair moved further than 1 metre apart for longer than 1 second, the contact recording ended. Each collared possum had a unique identification

number, the collar ID, in addition to their ear-tag ID. During subsequent trappings, loggers were checked to ensure proper functionality, malfunctioning loggers replaced, and the data downloaded. The timeline of these trapping sessions is illustrated in Fig. 2.2.

During later trapping sessions, possums were anaesthetised, and inspected for existing TB infection. A subgroup of the collared possums were injected subcutaneously with *Mycobacterium bovis*, the causative agent of bovine tuberculosis. The populations were then monitored for secondary infections.

In Season 1, four possums on each site were experimentally infected in July 2012. However, on sites A and B, seven of the eight possums unexpectedly died in the month following infection due to cold and wet weather, and another was unable to be found again. This necessitated the infection of four extra individuals on both sites in November 2012. Individuals in Season 1 were chosen dependent on their number of neighbours, the number of possums they had made contact with up to that point. On sites A and C, possums designated as 'highly-social', meaning that their number of neighbours were in the top 25% of those on the site, were chosen. Conversely, possums on sites B and D were chosen from 'low social' individuals, those whose numbers of neighbours were in the bottom 25%. In Seasons 2 and 3, twice as many individuals were infected, however only two sites were used for each season (sites C and D in Season 2, and sites A and B in Season 3). Unlike in Season 1, the individuals experimentally infected in Seasons 2 and 3 were not selected based on the number of possums they made contact with. An illustration of the experimental infections, and subsequent secondary infections detected is shown in Fig. 2.3.

Between Seasons 1 and 2, any possums trapped on sites C and D in January 2013, and any trapped on sites A and B in April 2013 were euthanised. This process is henceforth referred to as depopulation. The depopulation did not affect every individual trapped during the entirety of Season 1, so the populations in Seasons 2 and 3 are a combination of possums from Season 1 and new possums. Following euthanisation, a post-mortem examination was conducted to determine the TB status of each possum.

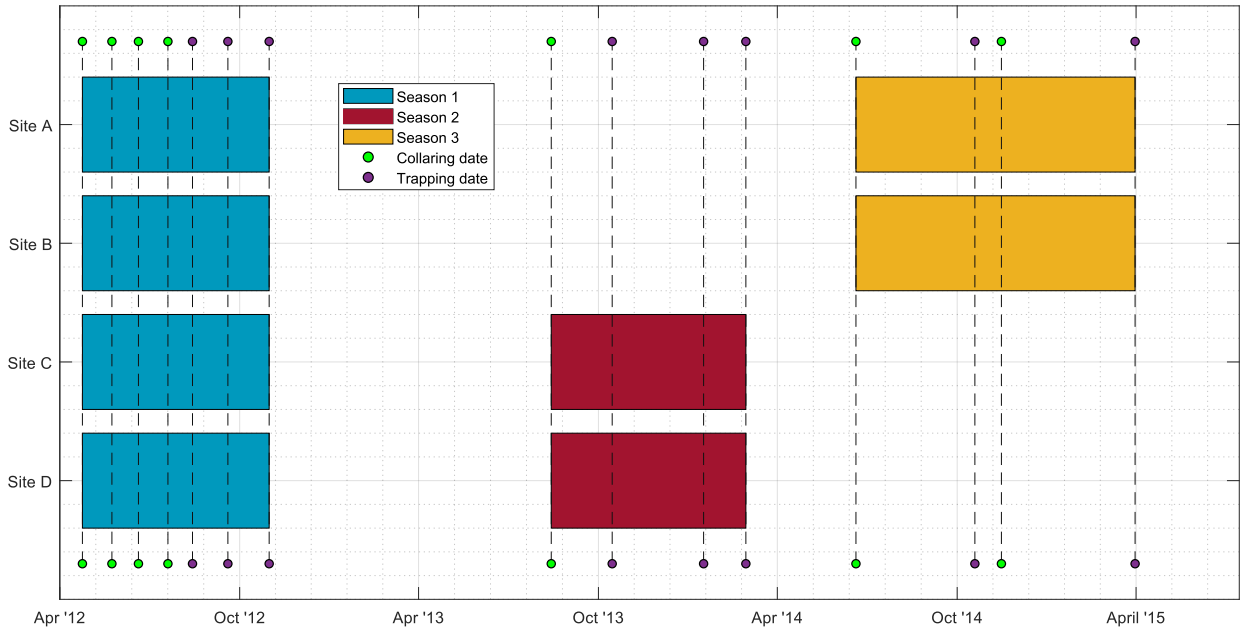



Figure 2.2: **An illustrated time-line of the trapping and collaring of possum populations.** The shaded areas indicate when individuals on that site were collared, with the three different colours referring to the three different seasons. The green and purple dots indicate weeks during which trapping took place, with the green dots indicating that collaring as well as trapping took place.

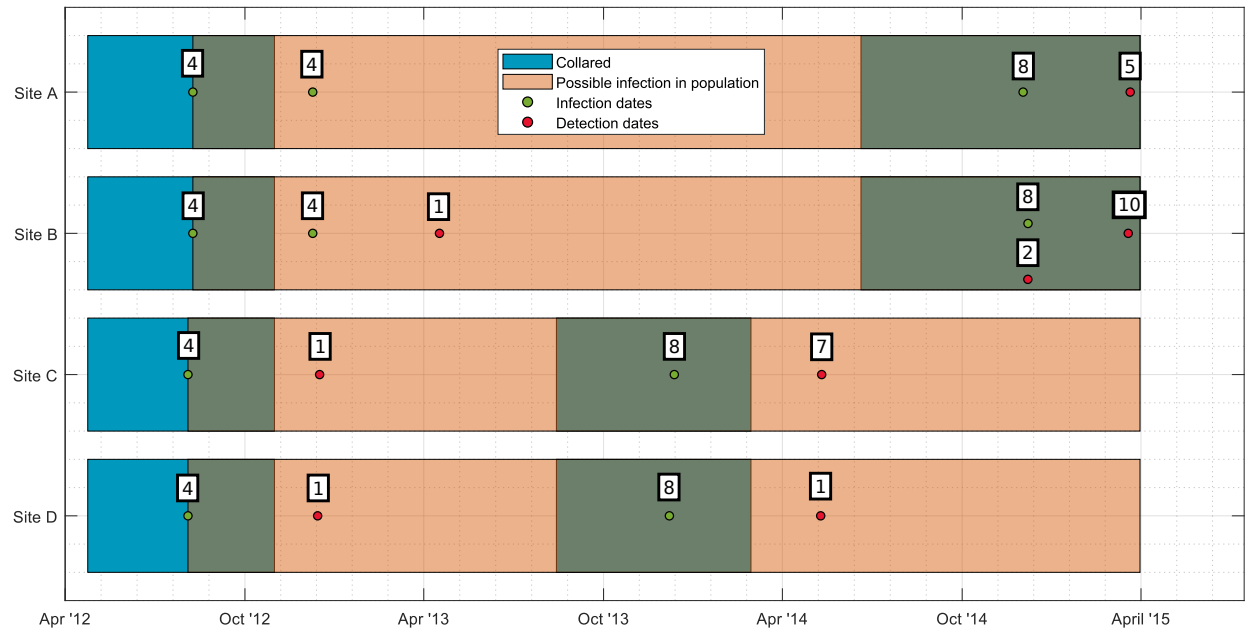



Figure 2.3: **An illustration of the time-line of the observation of the transmission of bovine tuberculosis in the possum populations.** The blue shaded areas indicate when individuals on that site were collared. The orange areas indicate the period of the study when deliberately infected individuals were in the population. The green dots indicate the dates where the deliberate infections took place, with the numbers above the dots indicating the number of individuals who were deliberately infected. The red dots indicate dates where secondary infections were detected, with the numbers above the dots indicating the number of secondary infections detected.

## 2.2 Data of contacts between common brushtail possums in the Orongorongo Valley

In total, across the three seasons, the data set consists of 127583 records. Each record in the data set details one contact between a pair of collared possums, and contains the collar IDs of both possums, the date and time the contact began recording, and the length of time, in seconds, of the contact recording.

### 2.2.1 Data cleaning

Of the 127583 records, 7510 were corrupted or invalid. Invalid records included those referring to contacts that occurred outside of the date range during which the collars could be guaranteed to be working (2122), and self contacts (20). There were 120073 usable records (114929 in Season 1, 617 in Season 2, and 4527 in Season 3).

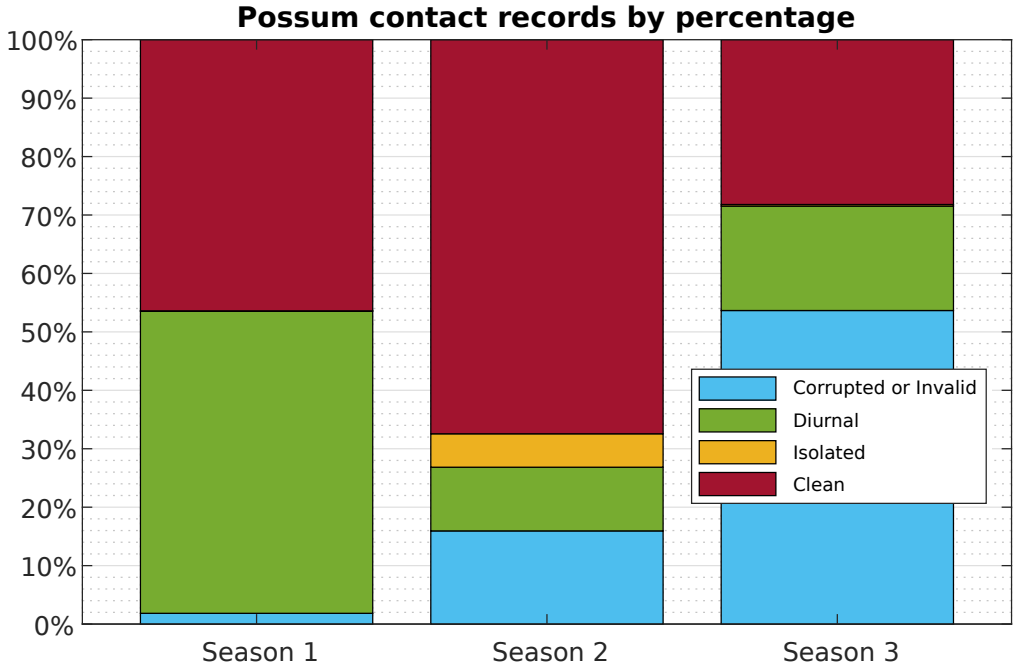
Data were cleaned by removing diurnal records (records recorded during the daytime), and then removing isolated records (one second long records that are the only record of a contact between a given pair). We document the effect of each step of the cleaning process on the data, by examining the remaining number of records and amount of contact time, as well as the structure of the contact networks produced. Empirical contact networks are produced by representing possums as nodes, and edges exist between a pair of possums if either possum recorded a contact between them during the study. The locations of the nodes are not representative of home-range centres or den locations, as we have no accompanying spatial information in our data to utilise. The contact networks are examined because data cleaning may result in removal of edges linking possum pairs. An illustration of the cleaning process on the contact networks from Season 2 is shown in Fig. 2.5.

| Season | Raw data | Corrupted or invalid | Usable data | Nocturnal records | Isolated nocturnal records | Clean data |
|---|---|---|---|---|---|---|
| 1 | 117081 | 2152 | 114929 | 54377 (47.31%) | 44 | 54333 (47.28%) |
| 2 | 734 | 117 | 617 | 537 (87.03%) | 42 | 495 (80.23%) |
| 3 | 9768 | 5241 | 4527 | 2782 (61.45%) | 27 | 2755 (60.86%) |

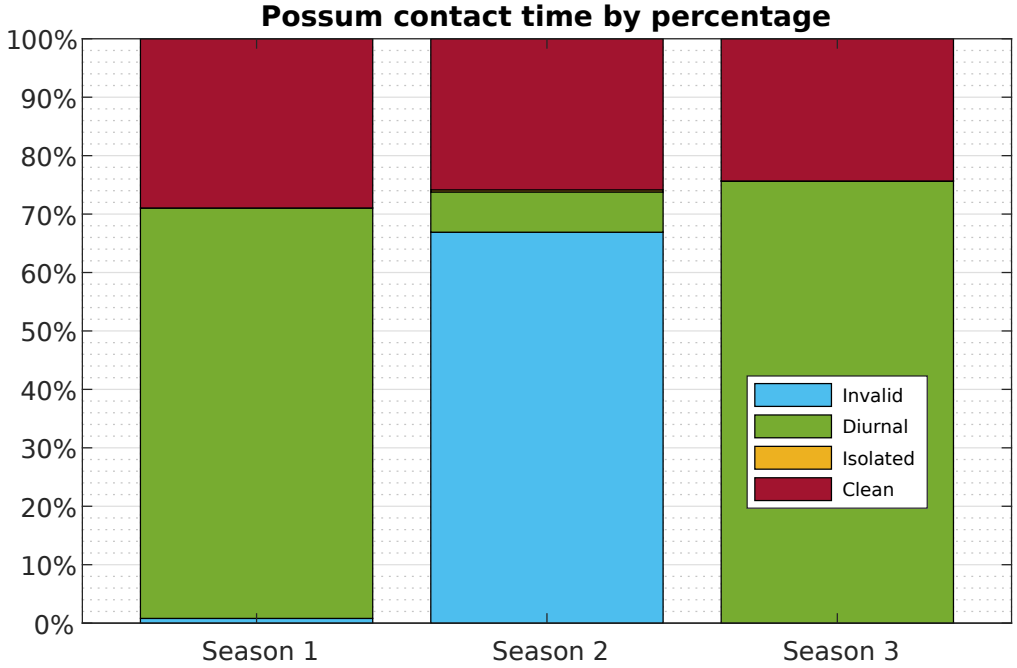
Table 2.1: **Breakdown of records at each stage of cleaning.** The effect of cleaning on the number of records varies between seasons, with Season 1 most affected (having only 47.28% of its records remaining, and Season 2 least affected (with 80.23%).

| Season | Raw time | Invalid | Usable time | Nocturnal time | Isolated nocturnal records | Clean data |
|---|---|---|---|---|---|---|
| 1 | 24404498 | 196709 | 24207789 | 7066552 (29.19%) | 44 | 7066508 (29.19%) |
| 2 | 12092 | 8089 | 4003 | 3169 (79.17%) | 42 | 3127 (78.12%) |
| 3 | 2155991 | 7 | 2155984 | 524837 (24.34%) | 27 | 524810 (24.34%) |

Table 2.2: **Breakdown of total time, in seconds, summed across all recorded contacts, at each stage of cleaning.**




(A) Breakdown of the effect of the different stages of cleaning on the possum contact records.




(B) Breakdown of the effect of the different stages of cleaning on the possum contact time.

Figure 2.4: **The percentage of remaining records after data cleaning is approximately the same across seasons. The percentage of contact time after data cleaning varies between seasons.**

### Removing diurnal records

Our initial models of possum movement and interaction only model nocturnal behaviour, when possums have left their dens and are moving around in their environment [Cowan and Clout, 2000]. As such, only nocturnal records are appropriate when verifying these models. Therefore, we remove records recorded during the daytime. The presence of daytime records may indicate that den-sharing has occurred, as it is otherwise unlikely that two possums would meet during the daytime. Simultaneous sharing of dens between possums does occur [Fairweather et al., 1987], as may sequential den-sharing [Green and Coleman, 1987].

We ignore seasonal variations in times of sunset and sunrise, and define diurnal records as records which began recording between 6am and 5:59:59pm. This assumption was made for two reasons. The first reason is pragmatism. There are algorithms for calculating the times of sunrise and sunset at a particular location on a given day [Meeus, 1998]. However, implementing these algorithms requires an understanding of relevant astronomical concepts that we do not possess and cannot justify allotting time and effort to gaining possession of. The second reason is due to assumptions made in the individual-based model that we discuss in Chapter 3. For ease of calculations and parameter estimation, a fixed 12 hour long nocturnal period is assumed. Therefore, it is necessary to make an analogous assumption when cleaning the data. By disregarding seasonal variations in the times of sunset and sunrise, a potentially significant amount of data could be erroneously included or excluded. This could potentially result in the exclusion of viable contact data or inclusion of inviable data. The latter refers to records labelled nocturnal in our cleaning, but were actually diurnal in reality.

The effect of removing diurnal records on the remaining number of records and contact time varies between seasons, and is summarised in Table 2.1 and Table 2.2. Season 2 is the least affected, with 87.03% of records and 79.17% of recorded contact time remaining. With regards to the number of records, Season 2 is followed by Season 3 and then by Season 1, with 61.45% and 47.31% remaining respectively. However, the opposite is true of the recorded contact time, with 29.19% and 24.34% of contact time in Seasons 1 and 3 remaining respectively.

By examining the contact networks after diurnal records records are removed, the effect of this stage of cleaning is shown to be minimal. The number of unconnected individuals is not increased in Seasons 1 and 2, and only increased by one in Season 3. Season 1 is the most affected with respect to the removal of edges,

with 98.18% of the edges remaining, on average across the four sites. In contrast, 98.51% of edges remain in the networks of Season 2, and 99.05% of edges in Season 3's networks. This is further summarised and expanded upon in Table 2.3. The similar effect on each season's contact networks indicates that a similar proportion of possum pairs were den-sharing.

Whilst a high percentage of recorded contact time in Seasons 1 and 3 is diurnal, the corresponding percentage of diurnal records is smaller (illustrated in Fig. 2.4A and Fig. 2.4B). This indicates that, on average, the diurnal records had a larger contact length than the nocturnal records. Due to possums' solitary and mostly nocturnal movements [Day et al., 2000], this gives an indication that there was den-sharing occurring during these field seasons.

**Removing isolated records**

We define a record as being an *isolated record* if it is the sole record of any contact between a pair of possums (meaning it has no corresponding record in the other possum's record corpus) across the entire field season, and has a contact length of one second. A contact between a pair of possums lasting at least one second is sufficient for TB transmission [Rouco, C., personal communication, August 16, 2020]. Because an isolated record is the only reference in either possum's record corpus of a contact occurring, and only lasts a total of one second, we make the assumption that it does not represent a valid contact. Subsequently, we are assuming that any non-isolated records refer to contacts during which TB could have been transmitted.

There are 113 isolated records in total across the three seasons, with similar numbers between the seasons. Removing these records has the greatest affect on both the number of records and total contact time for Season 2, predominantly due to the relatively small number of records in the usable data. This is summarised in Tables 2.1 and 2.2. A similar effect is seen in the contact networks produced, summarised in Table 2.3. Season 2 sees the greatest percentage decrease in the number of edges, with Season 3 having the greatest increase in the number of unconnected individuals.

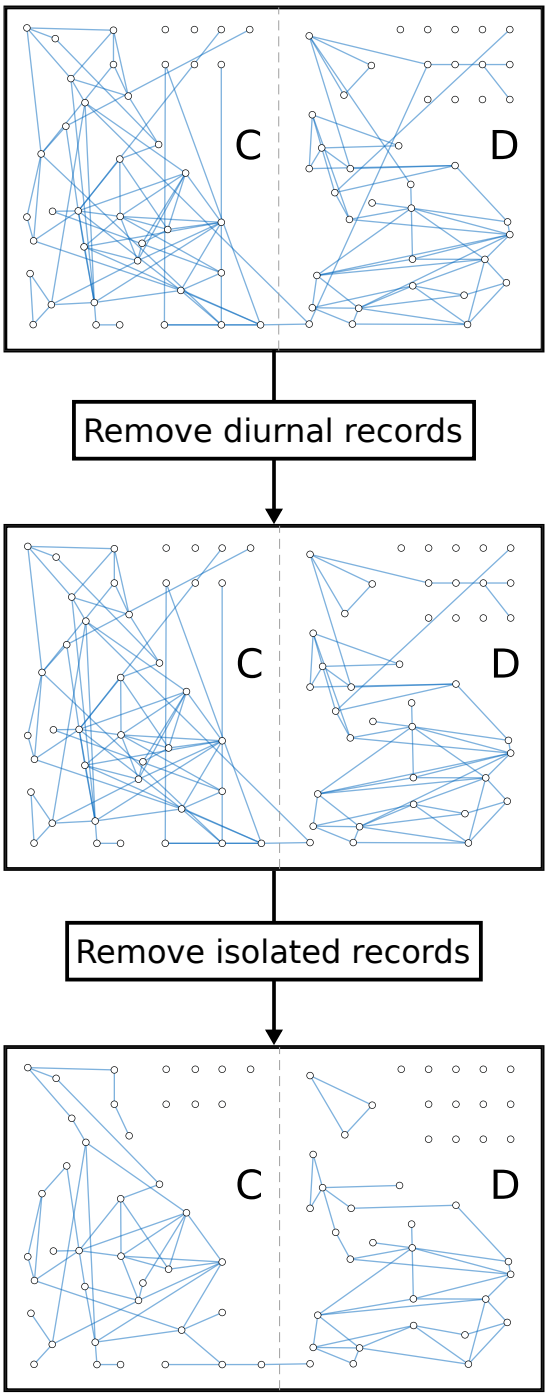



Figure 2.5: **Illustration of the effect of data cleaning on empirical contact networks representing possum populations in Season 2.** The effect of data cleaning is greatest when removing isolated records.

| | | | Usable data | | Nocturnal data | | Clean data | |
|---|---|---|---|---|---|---|---|---|
| Season | Site | Network Size | N. edges | Unc Inds* | N. edges | Unc Inds* | N. edges | Unc Inds* |
| 1 | A | 39 | 209 | 0 | 202 (96.65%) | 0 | 189 (90.43%) | 0 |
| 1 | B | 41 | 101 | 0 | 100 (99.01%) | 0 | 88 (87.13%) | 0 |
| 1 | C | 40 | 304 | 0 | 301 (99.01%) | 0 | 293 (96.38%) | 0 |
| 1 | D | 39 | 140 | 2 | 138 (98.57%) | 2 | 128 (91.43%) | 2 |
| 2 | C & D | 80 | 134 | 11 | 132 (98.51%) | 11 | 90 (67.16%) | 20 |
| 3 | A & B | 82 | 105 | 23 | 104 (99.05%) | 24 | 77 (73.33%) | 34 |

Table 2.3: **Table of summary values for the empirical contact networks produced from the totally cleaned data.** UncYou should define this shorthand in the Table captionThe number of edges of the empirical contact networks in is reduced slightly when the contact data is cleaned. The number of unconnected individuals is not increased by cleaning the data in Season 1, but is increased by 9 and 11 in Season 2 and 3 respectively. On average, the number of edges is reduced by 14.5% in Season 1, by 34.3% in Season 2, 31.1% in Season 3. Further, by removing isolated records, we remove a singular edge connecting sites C and D in Season 1, and two edges connecting sites C and D in Season 2.
*Shorthand for Unconnected Individuals.

### 2.2.2 Summary of cleaned data

After cleaning, the number of of remaining records is 57583, with 54333 from Season 1, 495 from Season 2, and 2755 from Season 3, representing 47.28%, 80.23%, and 60.86% of the usable records respectively. This indicates that the proportion of clean records is dependent on the season during which they were collected. A similar conclusion can be drawn for the amount of contact time recorded. After cleaning, a total of 7594445 seconds of recorded contact time remains, 7066508 from Season 1, 3127 from Season 2, and 524810 from Season 3. Of the usable contact time, Season 3 retains the least, with only 24.34% remaining, then Season 1 with 29.19%, and Season 2 the most, with 78.12%. Overall, this indicates that a larger proportion of the records, and contact time, between individuals in Season 2 occurred during nocturnal hours. However, they met far less frequently than in either of Seasons 1 and 3. The breakdown of the data into clean, diurnal and other categories is illustrated in Fig. 2.4A and Fig. 2.4B.

Data cleaning has a greater effect on Seasons 2 and 3 than Season 1. The number of edges is reduced by 8.66% in Season 1, 32.84% in Season 2, and 26.67% in Season 3. The number of unconnected individuals does not increase in Season 1, but increases by 9 and 11 in Seasons 2 and 3 respectively. The exact numbers are given in Table 2.3.

The variation between the field seasons can be explained by seasonality and the depopulations that occurred. The contact data for each field season was collected at slightly different times of the year. As such, the possum behaviour observed will vary field season to field season [Rouco, C., personal communication, August 16, 2020]. The populations in Season 1 were, presumably, stable. However, due to the depopulations taking place after Season 1, the contacts recorded for Seasons 2 and 3 represent social relationships that were recovering after being disrupted, and therefore the amount of data collected would have been impacted.

### 2.2.3 Analysis of cleaned data

From our cleaned data, for each pair of possums $i$ and $j$ who came into contact, we produce three results that measure the length and relative importance of their contacts:

1. The nightly contact probability $P_{ij}$, calculated as the proportion of nights, between the first and last contact (inclusive), during which a contact occurred.

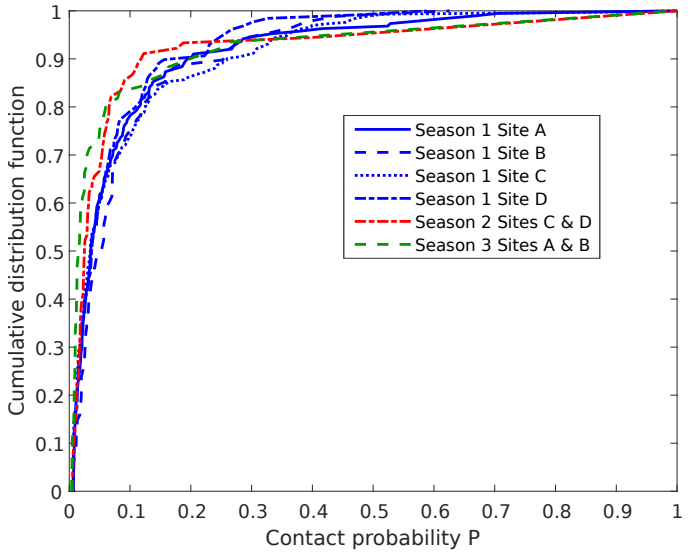
2. The average nightly first contact time $F_{ij}$. Each night during which a contact was made, the time the first record began recording is taken. The value of $F_{ij}$ is calculated as the average of these, over the number of nights during which a contact occurred.

3. The average total nocturnal time, $T_{ij}$, spent in contact. For each night where a contact occurred, the sum of all record lengths was taken. $T_{ij}$ is calculated as the average of these, over the number of nights during which a contact occurred.

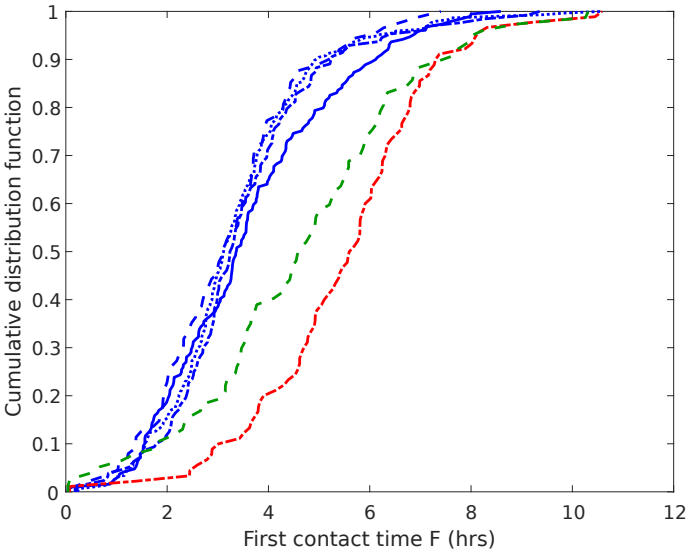
Here, a night is defined as the 12 hour period of time from 6pm to 5:59:59am. The empirical values of $P_{ij}$, $F_{ij}$ and $T_{ij}$ that we calculate will be used throughout this thesis. For example, in Chapter 3 we use values of $P_{ij}$ to estimate parameters of our possum movement model.

Mismatches between records for a single night are determined to have occurred either when only one possum in a pair recorded a contact, or when both possums recorded a contact but the the first record's time or the total of all contact lengths differ. Any night during which either possum in a pair recorded a contact is used when calculating $P_{ij}$. Instances where only one possum in a pair recorded a contact on a given night, the values from that night are used when calculating $F_{ij}$ or $T_{ij}$. If both possums in a pair recorded a contact on a given night, but the time of the first contact differs, the minimal value is taken, and if the total nocturnal contact times differ, an average is taken.

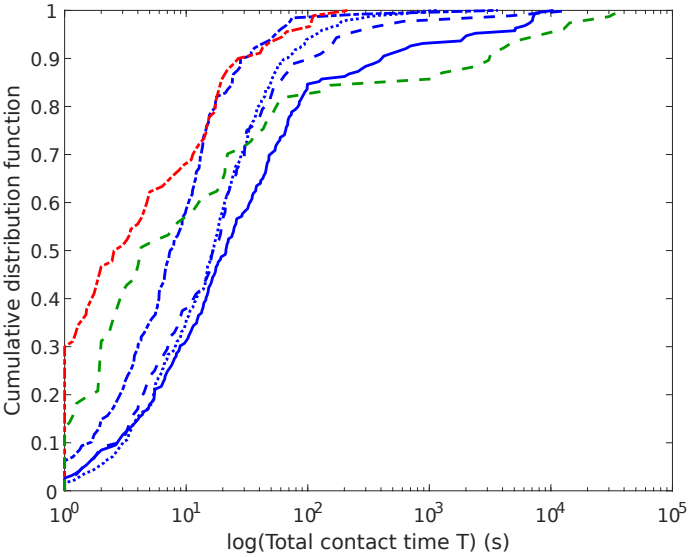
Cumulative distribution functions (CDFs) of $P_{ij}$, $F_{ij}$ and $T_{ij}$ are shown in Fig. 2.6. The CDFs of $P_{ij}$ vary little site to site when focusing on Season 1, as shown in Fig. 2.6A. However, comparing season to season, the proportion of smaller $P_{ij}$ values for Seasons 2 and 3 is greater than for Season 1. This indicates that the possum pairs in Season 1 are meeting on a higher proportion of nights than the possum pairs of Seasons 2 and 3.




(A)




(B)




(C)

Figure 2.6: **Cumulative distribution functions of outputs from cleaned data.**

Throughout Season 1, the CDFs of $F_{ij}$ (shown in Fig. 2.6B) are similar, but the maximal values of $F_{ij}$ are markedly different, with 7.4 hours and 10.6 hours for sites B and C respectively. This could indicate that the topography of each site has an effect on the time taken for the first contact to occur. When considering Seasons 2 and 3, the distributions of $F_{ij}$ consist of higher values. This indicates that possums were taking longer to meet than those in Season 1. This, like the seasonal variation of the record corpus, can be explained by both seasonality and the recovery and re-establishment of disrupted social relationships after depopulation.

The values of $T_{ij}$ vary greatly site to site and season to season, as shown by Fig. 2.6C. In Seasons 2 and 3, there are a significantly higher proportion of one second values than in Season 1. In addition, the maximal value of $T_{ij}$ varies. In Season 2 it is 215 seconds, or approximately 5.5 minutes, whereas the maximal $T_{ij}$ value for site D in Season 1 is approximately 10.5 hours.

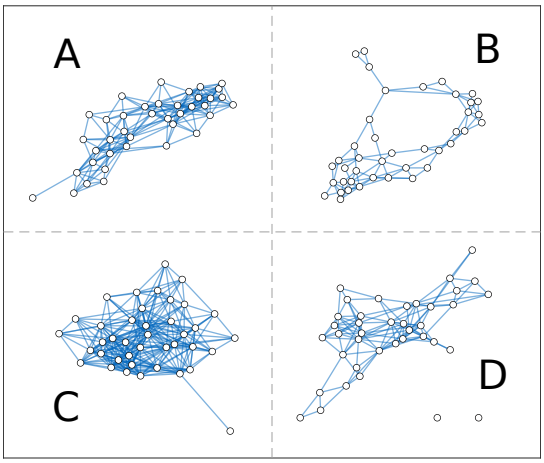
By also looking again at the distributions of $P_{ij}$, we can conclude that not only were possums in Season 2 meeting less often, but also for a shorter period of time. When comparing Season 1 to Season 3, we can conclude that possum pairs in Season 3 made contact less often, but for a longer period of time. This is corroborated by examining the lengths of the records. As summarised in Table 2.4, the modal record length is the same across each season, at one second, whereas the median is the same in Seasons 1 and 2, but comparatively larger in Season 3. The mean record length is smallest at 6.32 seconds in Season 2, and largest in Season 3 at 190.49 seconds.

| | Cleaned data | | Record length (s) | | |
|---|---|---|---|---|---|
| Season | Time (s) | # Records | Mean | Median | Mode |
| 1 | 7066464 | 54292 | 130.06 | 1 | 1 |
| 2 | 9199 | 488 | 6.32 | 1 | 1 |
| 3 | 765577 | 2747 | 190.49 | 13 | 1 |

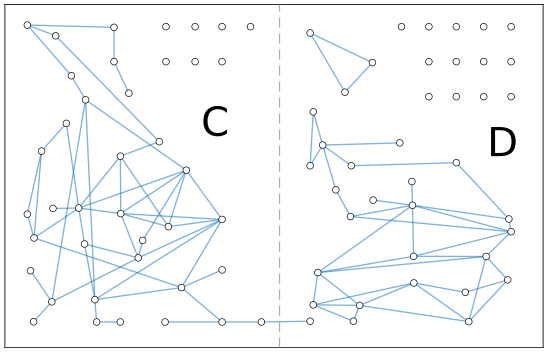
Table 2.4: **The distributions of record lengths vary between seasons.** The modal record length in each season is one second. The median, whilst the same in Seasons 1 and 2, but comparatively larger in Season 3. The mean record length is smallest at 6.32s in Season 2, and largest in Season 3 at 190.49s.

We produce binary contact networks from our cleaned data, as shown in Fig. 2.7. Fig. 2.8 shows the degree distributions of the empirical contact networks shown in Fig. 2.7. The distributions differ site to site, and a hypothesis for the variation between seasons is that possum population density (Table 3.1) is highest on site B, and lowest on site C. According to Richardson (2016), mean possum population density is significantly

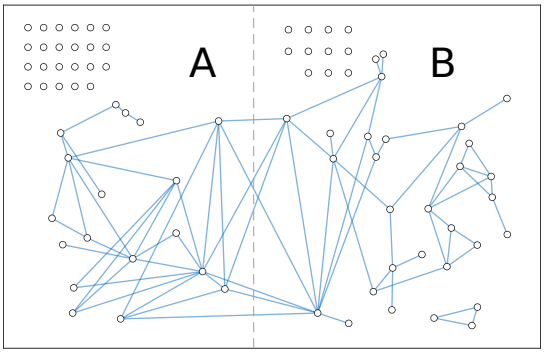
negatively correlated with mean home-range size. Individuals on site C will therefore have a larger home-range than those on site B, and make contact with a greater number of individuals. However, they would come into contact less frequently, resulting in lower values of the total contact time $T_{ij}$. Overall, this would lead to a higher mean degree, and could lead to the degree distribution we see for site C, in addition to the greater proportion of lower $T_{ij}$ seen in Fig. 2.6C.




(A) Empirical contact networks representing the populations in Season 1.




(B) Empirical contact networks representing the populations in Season 2.




(C) Empirical contact networks representing the populations in Season 3.

Figure 2.7: **Contact networks for each of the field seasons, produced from cleaned data.**

Fig. 2.9 shows the distribution of the numbers of nightly contacts. For each possum pair on each night during which a contact was recorded, the number of contacts between them were recorded. From this we can conclude that, for each site and season, the modal number of contacts per night, where a contact occurred, is one.

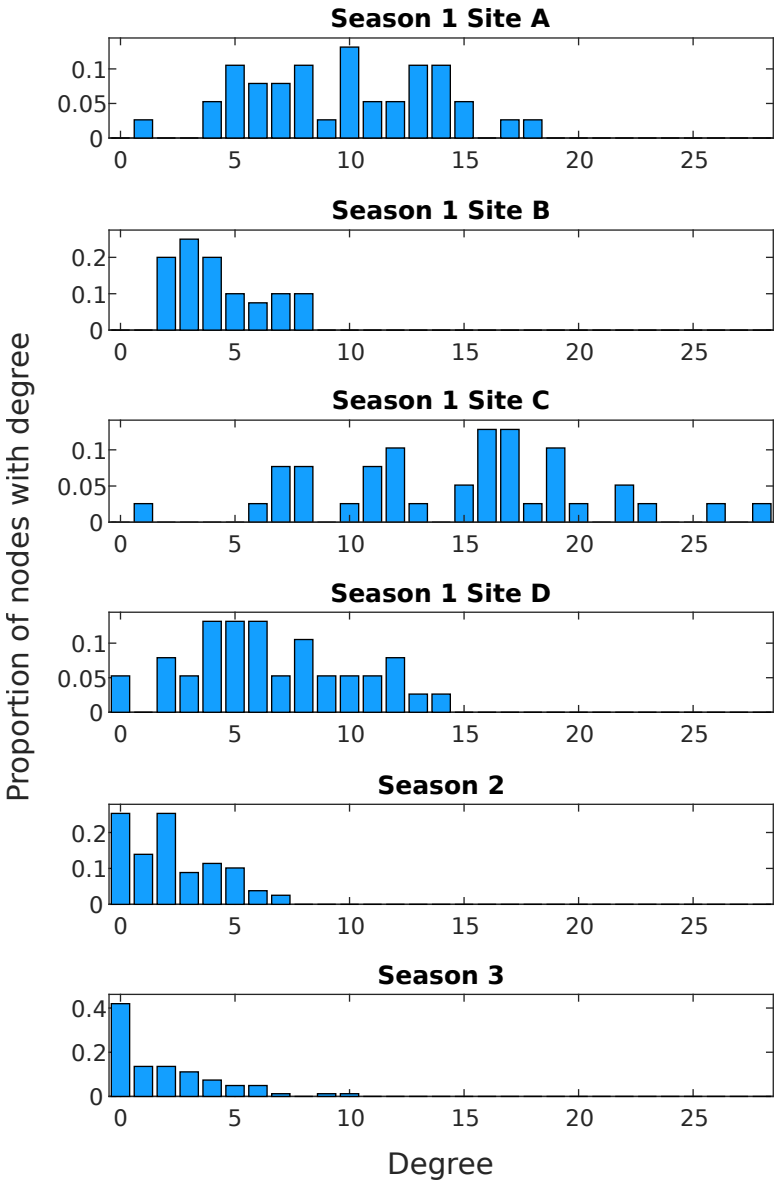



Figure 2.8: **The degree distributions of the empirical contact networks differ site to site.** The distributions differ site to site and season to season. An explanation for the variation between seasons is that possum population density (Table 3.1) is highest on site B, and lowest on site C. Richardson (2016) found that mean possum population density is significantly negatively correlated with mean home-range size. Therefore, individuals on site C for example will have a larger home-range than those on site B, and could make contact with more individuals. This would lead to a higher mean degree, and could lead to the degree distribution we see for site C. However, individuals on site C would come into contact less frequently, resulting in the greater proportion of lower $T_{ij}$ seen in Fig. 2.6C.

A table of summary statistics for the binary contact networks is shown in Table 2.5. From the table we can see that there is a positive correlation between the mean degree and the number of length-3-cycles. This could simply be a natural feature of networks, as the higher the mean degree, the greater number of edges, which would result in a greater number of length-3-cycles.

| Season | Site | Network Size | N. Edges | Unconnected Individuals | Mean Degree | Network density | Length-3-Cycles |
|---|---|---|---|---|---|---|---|
| 1 | A | 39 | 189 | 0 | 9.69 | 0.255 | 359 |
| 1 | B | 41 | 88 | 0 | 4.29 | 0.107 | 50 |
| 1 | C | 40 | 293 | 0 | 14.65 | 0.376 | 943 |
| 1 | D | 39 | 128 | 2 | 6.56 | 0.173 | 167 |
| 2 | C & D | 80 | 90 | 20 | 2.25 | 0.028 | 23 |
| 3 | A & B | 82 | 77 | 34 | 1.88 | 0.023 | 34 |

Table 2.5: **Summary statistics of the empirical possum contact networks produced from the cleaned data.** There is a positive correlation between the mean degree and the number of length-3-cycles.

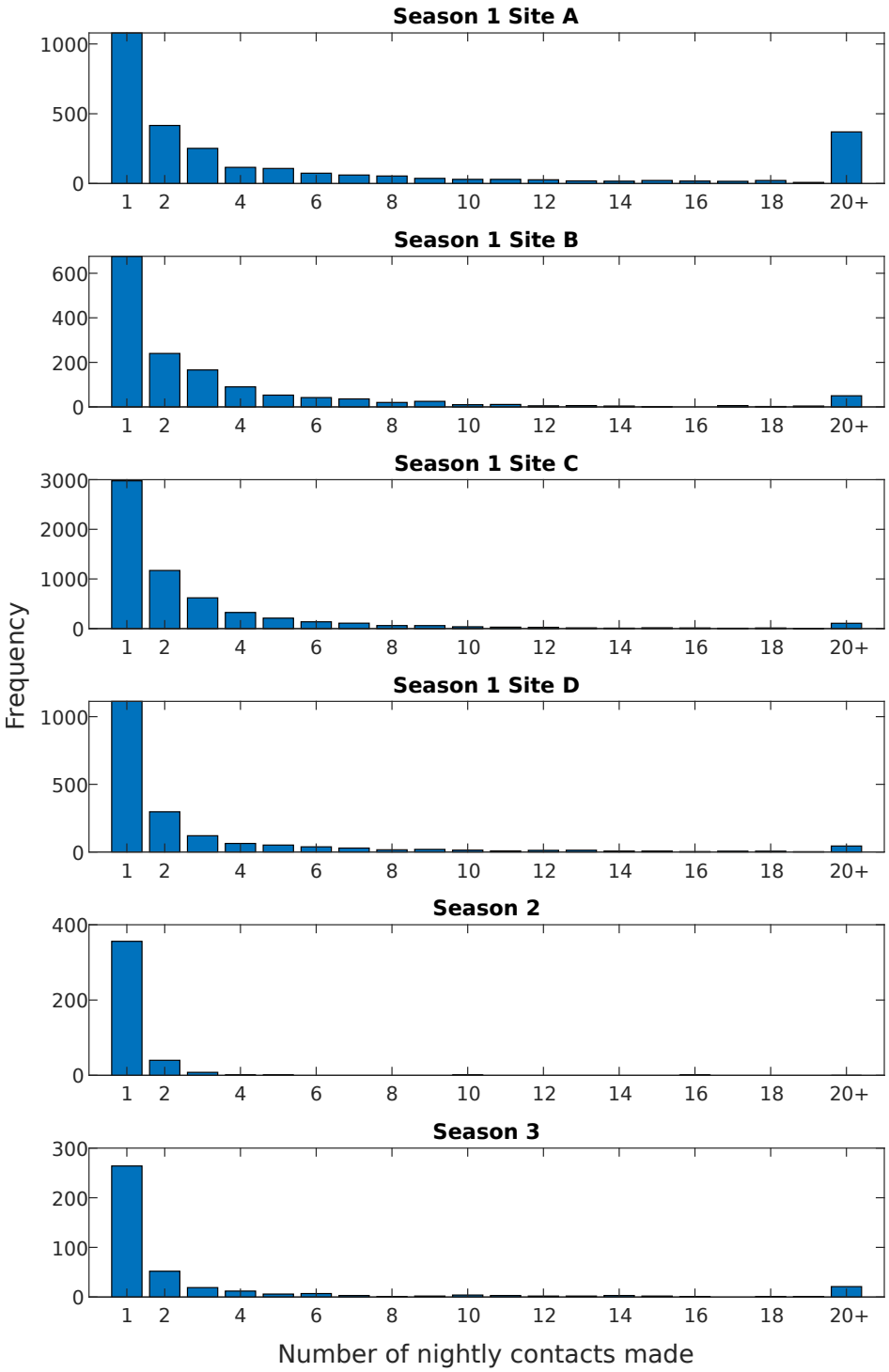



Figure 2.9: **The distribution of the numbers of nightly contacts.** For each possum pair on each night during which a contact was recorded, the number of contacts between them were recorded. As shown by the bar chart, the modal value for each site and season is one contact per night.

## 2.3 Data of the transmission of Bovine Tuberculosis between common brushtail possums in the Orongorongo Valley

Each record in the data details a single instance of the detection of a secondary infection, an infection found in an individual who was not in the seed population, and therefore not experimentally infected. Each record contains the unique ear tag designation(s), the first trapping date, site, and infection detection date. Records referring to collared possums also contained their collar ID. A summary of this data is given in Table 2.6.

There is no correlation in the data between the number of individuals that are experimentally infected, and the number counted as secondary infections. However, it should be noted that because of the time-span and the small size of the data, it is entirely possible that bovine tuberculosis spread to other possums before and after the populations were examined for disease. Fig. 2.10 shows the contact networks of the populations in each site and season, highlighting in green individuals who were primary infections, and those who were secondary infections in red. The two secondary infections of site B in Season 3, that were detected in November 2014, were as a result of infections conducted during Season 1. These individuals are highlighted in the blue box in the plot of site B in Fig. 2.10C.

We refer to a secondary infection as *explainable* if there exists a determinable infection pathway between the secondary infection and at least one primary infection. A determinable infection pathway is a path between two nodes in the network, where each node in the path is infected (either a primary of secondary infection). Examples of explainable secondary infections are given by the three red circular nodes in the plot of site C in Fig. 2.10B.

A secondary infection is *unexplainable* if there is no determinable infection pathway between any primary infection and the secondary infection. We can examine the unexplainable secondary infections, and categorise them according to the possible ways in which they may arise. They can arise in three possible ways:

(i) The individual is uncollared.

(ii) The individual is collared, but is not connected to a component of the network containing a primary infection.

(iii) The individual is collared and connected to a component containing a primary infection, but there are

| Season | Site | Experimental infection month | Infection detection month | Count |
|---|---|---|---|---|
| 1 | A | Jul-12 | - | 4 |
| | | Nov-12 | - | 4 |
| | B | Jul-12 | - | 4 |
| | | Nov-12 | - | 4 |
| | | - | Apr-13 | 1 |
| | C | Jul-12 | - | 4 |
| | | - | Dec-12 | 1 |
| | D | Jul-12 | - | 4 |
| | | - | Dec-12 | 1 |
| 2 | C | Nov-13 | - | 4 |
| | | - | Apr-14 | 1 |
| | | - | May-14 | 6 |
| | D | Nov-13 | - | 8 |
| | | - | May-14 | 1 |
| 3 | A | Nov-14 | - | 8 |
| | | - | Mar-15 | 5 |
| | B | Nov-14 | - | 8 |
| | | - | Nov-14 | 2* |
| | | - | Mar-15 | 10 |

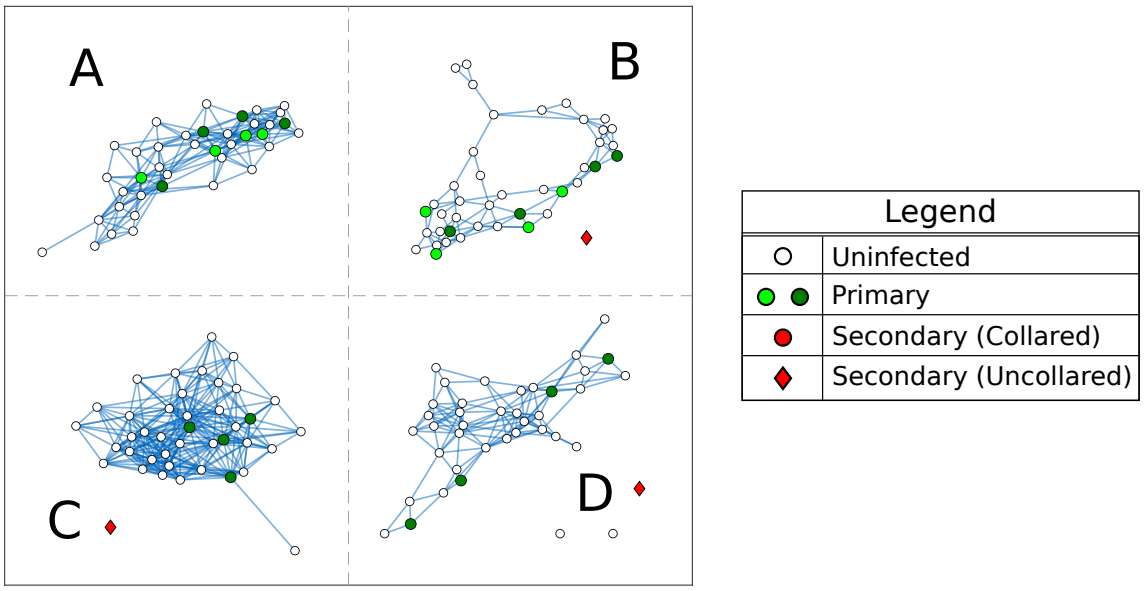
Table 2.6: **Table of experimental primary, and subsequent secondary, infections.** Dates individuals were experimentally infected on each site during each season, with counts of the number infected on each date. Months of secondary infection detections for each site and each season, with counts of the number of secondary infections detected during each month.
*These two individuals are the result of infections conducted during Season 1.

no determinable infection pathways containing the individual.

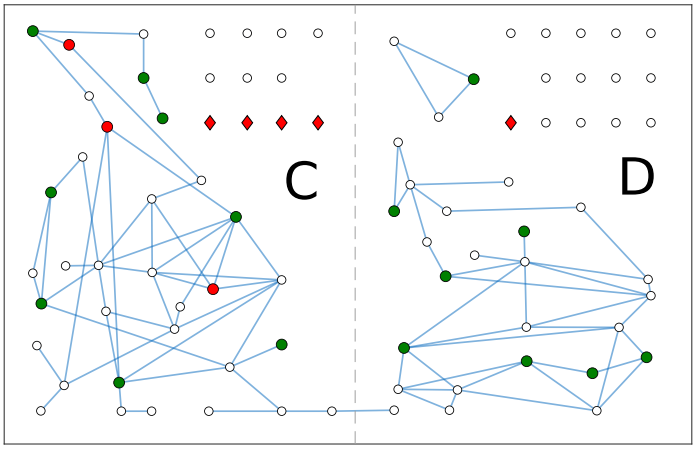
Examples of the first two types of unexplainable secondary infection are highlighted in Fig. 2.10C. The circular red node, highlighted in the blue box in Fig. 2.10C would have the singular example of a the third type of unexplainable secondary infection, however the individual's infection status was determined in November 2014, and is the result of deliberate infections carried out in Season 1. Therefore, it is in fact of the first type, as it was uncollared when infected. Table 2.7 gives the number of explainable and unexplainable secondary infections. The presence of unexplainable secondary infection highlights that there are individuals important to the spread of disease about whose interactions we know very little. This indicates the necessity for attempting to reconstruct the entire social contact network, given our small sub-population.

| | | Unexplainable secondary infections | | |
|---|---|---|---|---|
| Site | Explainable secondary infections | Type (i) | Type (ii) | Type (iii) |
| A | 0 | 0 | 0 | 0 |
| B | 0 | 1 | 0 | 0 |
| C | 0 | 1 | 0 | 0 |
| D | 0 | 1 | 0 | 0 |
| C | 3 | 4 | 0 | 0 |
| D | 0 | 1 | 0 | 0 |
| A | 0 | 5 | 0 | 0 |
| B | 0 | 10 | 1 | 1 |

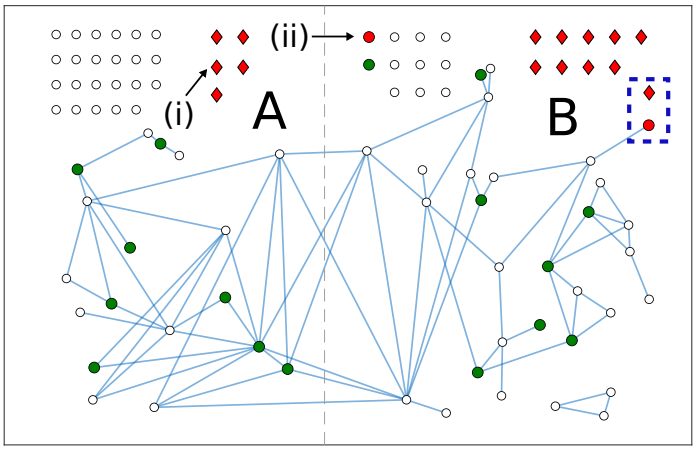
Table 2.7: **Numbers of explainable and unexplainable secondary infections.**




(A) Empirical contact networks representing the populations in Season 1.




(B) Empirical contact networks representing the populations in Season 2.




(C) Empirical contact networks representing the populations in Season 3.

Figure 2.10: **Binary contact networks of common brushtail possum populations, during the three seasons of the experiment.** The vertices represent individual possums (circular for collared and rhomboid for uncollared), with edges between possums pairs who had at least one contact during the respective field season. The green vertices represent experimental infections (In Season 1, July '12 infections are shown in light green and November '12 infections in dark green). The red vertices are secondary infections, those individuals trapped during later trapping sessions and found to be infected with TB. The experimental infection, as well as the secondary infection detection dates are shown in Table 2.6. The two nodes in Fig. 2.10C highlighted by the blue box were a result of infections conducted during Season 1. The other nodes highlighted in are examples of different types of unexplainable secondary infections.

# Chapter 3

# Using individual contacts to make weighted networks

## 3.1 Background

The preceding chapter presented and examined a data set of contacts between common brushtail possums in four distinct populations. In addition, metrics were calculated to describe the length and importance of contacts between possum pairs. These data have no spatial component, meaning there is no information about the locations of contacts, nor of where individuals den. As a result, whilst insight into possum contacts can be made from these data, individual movements are unable to be directly analysed.

In this chapter, we describe a model of possum movement and concurrent contacts, that will produce results about the length and importance of these contacts with an additional associated spatial component. Subsequently, this will also allow us to produce binary and weighted contact networks representing possum populations.

The movements of individuals in a possum population are modelled using a spatial, individual-based approach, specifically an unbiased random walk on a lattice structure. The choice of a random walk was made because of an assumption about the nightly activity or home-range of an individual. Richardson et al. (2017) provided parameter estimates for the population we study, that allow us to calculate home-range area estimations using the technique of Efford (2004) and Efford et al. (2005). However, they used a minimum convex polygon

approach, using trapping-location data, to estimate home-range sizes. We instead use the assumption of Efford et al. (2005), that the distribution of an individual's locations in their home-range can be described using a bivariate normal distribution. This assumption fits with results of individual location distributions produced by a random walk.

The modelling of possum movements using a random walk is not novel to this thesis. James et al. (2017) described the movements of possums using a mean-reversion random walk. Here, possums' movements tend towards the centre of their home-range, the area where they live, feed, and breed. As the distance from the centre increases, movements towards the home-range centre are more likely. This differs to the random walk we utilise, as there is no preference towards a particular location. The choice of an unbiased random walk was made because of the intention to investigate the effect of distance between dens on the contacts between possums, without the added complexity of home-range centre preference.

The movement and contact model produces simulated values of three measures of contact length and relative importance. Each of these measures is a function of the distance between two individuals' home-range centres. This allows a relationship between spatial information and contact results to be formed. We also investigate the effect of the distribution of home-range centres on results produced.

We model the structure of a population using a network, consisting of points representing individuals in a population, and a set of links between the points, representing relationships between individuals. Empirical contact networks created from data sets that, like ours are generated using proximity loggers, have been produced for animal populations including Tasmanian devils (*Sarcophilus harrisii*), in Hamede et al. (2009).

Though empirical contact networks of common brushtail possums have been produced before, not all have been produced through the use of explicit contact data. In Corner et al. (2003), an edge exists between a pair of possums if they share the same den at any point during the diurnal hours of the study. In addition to the edge condition, Corner et al. (2003) also differs in the respect that the possums were captive, meaning that their social structure is not necessarily the same as the free-living possums we study here. Porphyre et al. (2008), on the other hand, produce networks of free-living possums. However, whilst edges in our empirical contact networks are contingent on explicit contact, edges in the networks of Porphyre et al. (2008) exist between pairs if either sequential trap use or trapping in the activity range of another possum was

documented.

Empirical contact networks have previously been produced from data on the explicit contacts between possums. Ji et al. (2005) used data collected by proximity loggers to produce contact networks of wild-living common brushtail possums. Even the data we study has been used in this manner before. Both James et al. (2017) and Rouco et al. (2018) use the same data set as is analysed and used in this thesis, to produce contact networks, where edges exist between possum pairs if they came into contact during the field study.

After showcasing our model, and the results we can produce from it, we discuss three models to describe the home-range centre distribution of a subset of collared individuals from a larger population of trapped individuals. Following on from that, we describe our process for estimating the diffusion coefficient of our possums' random walks. The diffusion coefficient is a parameter that describes the rate at which individuals spread out through an area. Using these diffusion coefficients, we estimate the home-range size of our individuals, finding that home-range sizes calculated from data in Richardson et al. (2017) are well estimated by our model results on two field sites, and greatly overestimated on the other two field sites.

To validate our models, results are compared with the data set described in Section 2.2. We show that, with appropriate choice of parameters and of home-range distribution models, certain aspects of possum population structure and behaviour are replicable well to some extent by our work. The results from this chapter, specifically the weighted contact networks, will be used in Chapter 5 to model the spread of a bovine tuberculosis epidemic through a possum population.

## 3.2 Setting

In this section we describe mathematical techniques that are necessary for the models of possum population structure, movement, and contact that we produce in this chapter.

### 3.2.1 Random walks

A *random walk* is a stochastic process that describes the movements of an individual, with initial position vector $\mathbf{x}_0$, over a period of $n$ steps. The vectors $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ describing the individual's movements are a set of random vectors, each vector drawn from a distribution with a given probability mass function. The sequence of position vectors $(\mathbf{x}_k)_{k\geq 0}$, where $\mathbf{x}_n = \mathbf{x}_0 + \sum_{k=1}^{n} \mathbf{X}_k$, forms a random walk.

For a simple, unbiased random walk on a regular square lattice in $\mathbb{R}^2$, with spacing $\delta$, the random vectors $\mathbf{X}_k$ have a probability $1-p$ of being $(0,0)$ and probability $\frac{p}{4}$ of being one of $\{(\delta,0),(-\delta,0),(0,\delta),(0,-\delta)\}$. Each of the $\mathbf{X}_k$ are i.i.d. random variables, with expected value $(0,0)$ and covariance matrix $\frac{p\delta^2}{2}I_2$. Therefore, by the multidimensional central limit theorem, when $n$ is large the position vectors $\mathbf{x}_n$ are distributed according to a bivariate normal distribution, with mean $\mathbf{x}_0$ and covariance matrix $\frac{np\delta^2}{2}I_2$, i.e. $\mathbf{x}_n \sim N_2(\mathbf{x}_0, \frac{np\delta^2}{2}I_2)$. We define $t = n\tau$ and let $\mathbf{x}_t$ be the position vector of the individual at time $t$. This vector $\mathbf{x}_t$ is a random variable and, by taking the continuum limit $\frac{\delta}{\tau} \to \infty$, we have that $\mathbf{x}_t \sim N_2(\mathbf{x}_0, \mathcal{D}tI_2)$, where $\mathcal{D} = p\delta^2/2\tau$ is a constant, known as the diffusion coefficient [Codling et al., 2008].

Now let us consider two individuals, $i$ and $j$, who each follow unbiased random walks as described above, starting at their respective home-range centres $\mathbf{z}_i$ and $\mathbf{z}_j$. Let $\mathbf{z}_{ij} = (\mathbf{z}_i - \mathbf{z}_j)$ denote the vector difference between the home-range centres of $i$ and $j$. Then $\hat{d}_{ij} = \|\mathbf{z}_{ij}\|_2$ is the Euclidean distance between their home-range centres. Each individual's random walk has diffusion coefficient $\mathcal{D}$. We denote the vector different between $i$ and $j$ at step $n$ by $\mathbf{w}_n$, with $\mathbf{w}_0 = \mathbf{z}_{ij}$. The sequence $(\mathbf{w}_k)_{k\geq 0}$ of vector differences between the two individuals at each step also forms a random walk, and is expressed as $\mathbf{w}_n = \mathbf{z}_{ij} + \sum_{k=1}^{n} \mathbf{W}_k$, where the $\mathbf{W}_k$s

are a set of i.i.d. random vectors, each with an identical following probability mass function:

$$\mathrm{f}(\mathbf{w}) = \begin{cases} \frac{5p^2-8p+4}{4}, & \text{if } \mathbf{w} = [0,0], \\ \frac{p-p^2}{2}, & \text{if } \mathbf{w} \in \{[\delta,0],[-\delta,0],[0,\delta],[0,-\delta]\}, \\ \frac{p^2}{8}, & \text{if } \mathbf{w} \in \{[\delta,\delta],[\delta,-\delta],[-\delta,\delta],[-\delta,-\delta]\}, \\ \frac{p^2}{16}, & \text{if } \mathbf{w} \in \{[2\delta,0],[-2\delta,0],[0,2\delta],[0,-2\delta]\}. \end{cases}$$

From the central limit theorem, the vector difference between the two individuals at time $t$, $\mathbf{w}_t$, satisfies $\mathbf{w}_t \sim N_2(\mathbf{z}_{ij}, 2\mathcal{D}tI_2)$. The Euclidean distance between the individuals at time $t$ is $\|\mathbf{w}_t\|_2$, and we define $\Delta_t = \|\mathbf{w}_t\|_2/\sqrt{2\mathcal{D}t}$. In the continuum limit, we can express the variable $\Delta_t$ as the difference between two independent bivariate normally distributed random variables, both with covariance matrix $I_2$ and respective means $[0,0]$ and $[0, \hat{d}_{ij}/\sqrt{2\mathcal{D}t}]$. The square of the difference between two independent bivariate normals with unit covariance matrices and means $\mu_1$ and $\mu_2$ is distributed according to the noncentral chi-squared distribution $\chi^2(2, \mu_1^2 + \mu_2^2)$. Therefore, $D_t^2 \sim \chi^2(2, \hat{d}_{ij}^2/2\mathcal{D}t)$.

Let $P_t(\mathbf{w})$ be the probability of the random walk, consisting of the vector differences between two random walkers, being at $\mathbf{w}$ after $t$ steps, having started at the origin. From Montroll (1964), we have that

$$P_t(\mathbf{w}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} [\phi(\boldsymbol{\theta})]^t \exp\left(-i\boldsymbol{\theta} \cdot \mathbf{w}\right) d^2\boldsymbol{\theta}, \tag{3.1}$$

where $\phi(\boldsymbol{\theta}) = \sum_{\mathbf{w} \in \mathbf{V}} \mathrm{f}(\mathbf{w}) \exp\left(i\boldsymbol{\theta} \cdot \mathbf{w}\right)$ and $\mathbf{V} = \{[0,0],[\pm\delta,0],[0,\pm\delta],[\pm\delta,\pm\delta],[\pm\delta,\mp\delta],[\pm 2\delta,0],[0,\pm 2\delta]\}$.

Let $F_t(\mathbf{w})$ be the probability that the random walk, starting at the origin, reaches $\mathbf{w}$ for the first time after $t$ steps. Let $F(\mathbf{w}, s) = \sum_{n=1}^{\infty} s^n F_n(\mathbf{w})$ and $P(\mathbf{w}, s) = \sum_{n=1}^{\infty} s^n P_n(\mathbf{w})$. From Montroll and Weiss (1965), we have that

$$F(\mathbf{w}, s) = \sum_{n=1}^{\infty} s^n F_n(\mathbf{w}) = \frac{P(\mathbf{w}, s) - \delta_{\mathbf{w},\mathbf{0}}}{P(\mathbf{0}, s)} = \frac{(\sum_{n=1}^{\infty} s^n P_n(\mathbf{w})) - \delta_{\mathbf{w},\mathbf{0}}}{P(\mathbf{0}, s)}. \tag{3.2}$$

Here, in this context, $\delta_{\mathbf{w},\mathbf{0}}$ refers to the Kronecker delta function, and is unrelated to the previous and subsequent uses of $\delta$. Further, from Montroll and Weiss (1965) we also have the denominator of the fraction, $P(\mathbf{0}, s)$, is evaluated as:

$$P(\mathbf{0}, s) = \frac{1}{(2\pi)^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} [1 - s\phi(\boldsymbol{\theta})]^{-1} \, d^2\boldsymbol{\theta}. \tag{3.3}$$

By evaluating $P(\mathbf{0}, s)$, and equating coefficients of $s$, we can determine an expression for $F_n(\mathbf{w})$.

### 3.2.2 Population networks

A *binary network* $G$ is an ordered pair of sets $(V, E)$, where the elements of $V$ are called the *vertices* of $G$ and the elements of $E$ are called the *edges* of $G$. An edge of $G$ is an unordered double of the form $\{u, v\}$, for $u, v \in V$ (for ease of notation, we will denote the element $\{u, v\}$ by $uv$), and two vertices $u$ and $v$ are said to be *adjacent* in $G$ if there exists $uv \in E$. The *network density* of a network $G$ is the fraction of edges in the network out of the number of possible edges. The *degree* of a vertex $u$ is the number of edges incident to it. A triple of vertices forms a *3-cycle* if each pair in the triple are adjacent.

We can express a binary network using its adjacency matrix $A$, a symmetric $N \times N$ matrix with entry $a_{ij}$ equal to 1 if an edge exists between individuals $i$ and $j$, and 0 otherwise. By assigning to each edge a non-zero value, the edge weight, we form a *weighted network*. Similar to the binary network case, we can express this weighted network using a weighted adjacency matrix $W$, a symmetric matrix of size $N \times N$ where each non-zero entry $w_{uv}$ is the weight of the edge between $u$ and $v$. Individuals cannot contact themselves, so the diagonal entries of $W$ are zero, i.e. $w_{uu} = 0$.

### 3.2.3 Two-sample Kolmogorov–Smirnov test

The Two Sample Kolmogorov–Smirnov test is a statistical test used to compare two samples. For two samples, $A$ and $B$, the null hypothesis is that these two samples are both drawn from the same population or probability distribution [Lin et al., 2010]. The test is conducted by comparing the cumulative distribution functions of $A$ and $B$, $f_A$ and $f_B$ respectively, over the range of values in $A \cup B$, calculating the maximum absolute difference between $f_A$ and $f_B$. This difference is called the Kolmogorov–Smirnov test statistic and, for $A$ and $B$, is defined as:

$$\max_{z \in A \cup B} \|f_A(z) - f_B(z)\|_1 . \tag{3.4}$$

## 3.3 An individual-based model of movements and contacts

We assume there are $N$ individuals. During a 24-hour period, each individual spends a continuous period of 12 hours, corresponding to night-time, moving, and spends the other 12 hours denning in a single location, assumed to be the centre of their home-range. Seasonal variations in the times of sunset and sunrise mean that this assumption is not realistic, as the range in the length of night-time is approximately four hours, dependent on the location. Despite this, for the parameter estimation conducted later in this chapter, the assumption of a fixed night-time/day-time divide simplifies calculations considerably.

At the beginning of each night, individuals start from centre of their home-range. Let $\mathbf{z}_i$ be the centre of individual $i$'s home-range. Individuals move according to a simple unbiased random walk on a regular square lattice of spacing $\delta$, as per Section 3.2.1. We define an individual's home-range to be the area that, on average, 95% of their movements during any given night lie within. We assume that each individual's home-range is identical and circular. The assumption of identical sizes is made to simplify the calculations and analysis that will be performed later in this chapter. If instead the two individuals has different home-range sizes, then their random walks must also have differing diffusion coefficients. Further, if the home-ranges were not circular, then the covariance matrix of the distribution describing $\mathbf{x}_t$ would not be a scalar multiple of the identity matrix. This would prevent the use of the non-central chi-squared distribution, as unit covariance matrices are required.

The value of the diffusion coefficient $\mathcal{D}$ is assumed to be the same for each individual. A contact between a pair of individuals occurs if they occupy the same point on the lattice. Contacts have no effect on the random walk. After 12 hours, the individuals return directly to their home-range centre. No contacts occur during this return journey.

Fig. 3.1 shows the paths of two possums, who den 20 metres apart, over a period of 12 hours. Starting from their denning locations (green dots) each possum follows a simple unbiased random walk for a period of 12 hours, where each second they have a probability 0.6 of remaining in the same position, and probability 0.1 of moving one space away on the lattice. Their position after 12 hours is shown as red dots. Their random walks have a diffusion coefficient of $\mathcal{D} = 0.2\text{m}^2\text{s}^{-1}$. A contact occurs if the possums are less than 1 metre apart which, in this example, occurs 18 times over the 12 hours, and the locations of which are shown as white dots.

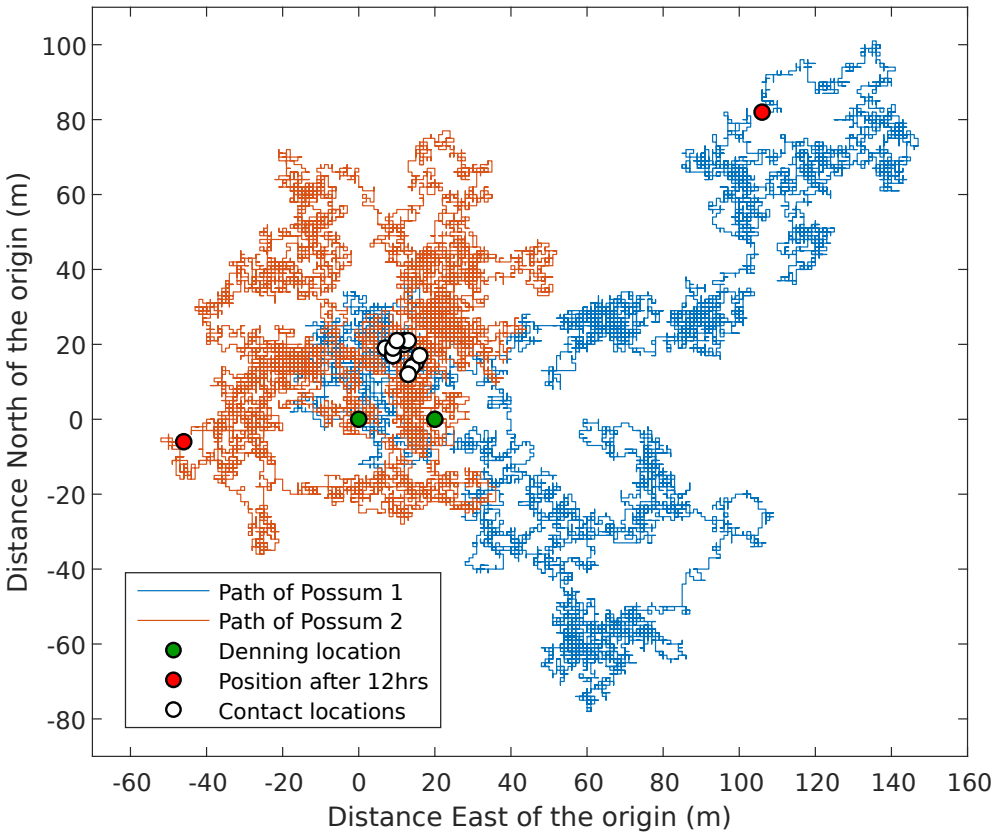



Figure 3.1: **An illustration of the random walks of two possums, who den 20 metres apart, over a period of 12 hours.** Each possum follows a simple unbiased random walk, where each second they have a probability 0.6 of remaining in the same position, and probability 0.1 of moving one space away on the lattice. As $p = 0.4$, $\delta = 1m$, and $\tau = 1s$, their random walks have a diffusion coefficient of $\mathcal{D} = 0.2\text{m}^2\text{s}^{-1}$. Starting from their denning locations (green dots) they walk for a period of 12 hours, after which they immediately return to their dens. Their positions after 12 hours are shown as red dots. A contact occurs if the possums are less than 1 metre apart, which occurs 18 times, and the locations of which are shown as white dots.

### 3.3.1 Measures of contact length and importance

In real-world possum populations, the distribution and length of possum contacts is not uniform. Some possum pairs may meet frequently, others may meet infrequently but for long periods of time, and others may never meet, among other scenarios. From our model, we produce three measures of the length and importance of contacts between a given pair of individuals for a given night:

1. The probability that at least one contact will occur.
2. The expected first time that a contact occurs, given that a contact does occur.
3. The expected total time spent in contact.

Each of these measures depends only on the distance $\hat{d}_{ij} = \|\mathbf{z}_{ij}\|_2 = \|\mathbf{z}_i - \mathbf{z}_j\|_2$ between the home-range centres of individuals $i$ and $j$, and on the diffusion coefficient $\mathcal{D}$ of their random walks. As per Section 3.2.1, we let $\mathbf{w}_t$ denote the vector difference between $i$ and $j$ at time $t$.

There exist results for 2-dimensional random walks that can be utilised to give analytical values of the first two measures [Montroll, 1964, Montroll and Weiss, 1965, Lindenberg et al., 1980]. In the following section, we give an example of how to apply results from Montroll (1964) and Montroll and Weiss (1965) to the first two contact measures, the probability that at least one contact will occur, and the expected first time that a contact occurs, given that a contact does occur. Unfortunately, an expression necessary for their evaluation has so far proven intractable to evaluate. Therefore, we use random walk simulations to generate our results. With the third measure, we use a standard analytical result from the continuum limit theory of random walks.

**Probability $P_{ij}$ of a contact occurring.**

We define $P_{ij}$ as the probability that at least one contact will occur between individuals $i$ and $j$ during a given night. This is equivalent, from the perspective of the real-world study, to the proportion of nights a particular possum pair come into contact with each other over an observation period. We can calculate an

analytical value for $P_{ij}$ as:

$$\begin{aligned} P_{ij} &= P(\exists t,\ 0 \leq t \leq t_{max},\ \text{s.t. } \mathbf{w}_t = \mathbf{0}) \\ &= 1 - P(\forall t,\ 0 \leq t \leq t_{max},\ \text{s.t. } \mathbf{w}_t \neq \mathbf{0}) \\ &= 1 - \prod_{t=0}^{t_{max}} P(\mathbf{w}_t \neq \mathbf{0}) \\ &= 1 - \prod_{t=0}^{t_{max}} (1 - P(\mathbf{w}_t = \mathbf{0})) \\ &= 1 - \prod_{t=0}^{t_{max}} (1 - P_t(\mathbf{z}_{ij}))\,, \end{aligned}$$

where $t_{max} = 12$ hours and $P_t(\mathbf{z}_{ij})$ is the probability of the random walk being at the point $\mathbf{z}_{ij}$ after $t$ steps, having started at the origin. The final step uses the symmetry of random walks to establish that $P(\mathbf{w}_t = \mathbf{0}) = P_t(\mathbf{z}_{ij})$. From Section 3.2.1, we have an expression for $P_t(\mathbf{z}_{ij})$ in the form of Eq. (3.1). However, the integral in Eq. (3.1) has so far proven intractable to evaluate, so instead we use random walk realisations to produce values of $P_{ij}$.

We calculate $P_{ij}$ for a particular pair of individuals, using realisations of two random walkers, on a square lattice of spacing $\delta = 1$ metre. These random walks are performed for the equivalent of 12 hours, with a time-step of $\tau = 1$ second. The values of $\delta$ and $\tau$ are chosen to correspond, respectively, with the contact distance and separation time of the proximity loggers used in the experiment described in Section 2.2.

The individuals start at $[0, 0]$ and $[x, 0]$, for $x$ an integer multiple of $\delta$. Both random walks have a diffusion coefficient $\mathcal{D} = p\delta^2/2\tau$. To alter the value of $\mathcal{D}$, the probability $p$ is altered. The probability $P_{ij}$ is calculated from simulations as the proportion during which a contact occurs. The values of the lattice spacing $\delta$ and time step $\tau$ chosen result in a maximal diffusion coefficient of 0.5 $\text{m}^2\text{s}^{-1}$. To increase the range the maximal value of $\mathcal{D}$, the time step $\tau$ is decreased.

Calculating $P_{ij}$ for a given value of $\mathcal{D}$ is computationally expensive. Because of the quantity of diffusion coefficients used in this chapter, we first calculate $P_{ij}$ for a fixed value of the diffusion coefficient, and then produce an approximation of $P_{ij}$ for a different diffusion coefficient $\mathcal{D}'$, by scaling the distance $x$ between home-range centres by $\sqrt{\frac{\mathcal{D}'}{\mathcal{D}}}$. The fixed value chosen was 0.2 $\text{m}^2\text{s}^{-1}$, and was chosen because it was of the same order of magnitude as the largest diffusion coefficient estimated in Section 3.5.

Figure 3.2A shows the contact probability $P_{ij}$, calculated for distances between 0 m and 500 m. For each distance, $P_{ij}$ was calculated by performing $10^6$ realisations of a pair of random walkers as described above. The solid lines show the values of $P_{ij}$, calculated for each distance by performing $10^6$ realisations of a pair of random walkers as described above, for diffusion coefficients 0.02 $\mathrm{m^2 s^{-1}}$, 0.1 $\mathrm{m^2 s^{-1}}$, and 0.2 $\mathrm{m^2 s^{-1}}$. The dashed lines show the approximated values of $P_{ij}$, calculated for $\mathcal{D}' = 0.1$ $\mathrm{m^2 s^{-1}}$ and $\mathcal{D}' = 0.02$ $\mathrm{m^2 s^{-1}}$, using a fixed value of $\mathcal{D} = 0.2\mathrm{m^2 s^{-1}}$. As the distance between the centres of the individuals' home-ranges increases, the contact probability decreases. Conversely, for any given initial distance, $P_{ij}$ increases as the diffusion coefficient increases. In practice, as the distance between two individuals' home-range centres is not necessarily an integer multiple of $\delta$, we obtain $P_{ij}$ by performing realisations of random walks for a range of values of $x$, and then interpolate from the resultant probabilities.

**Expected first contact time $F_{ij}$.**

We define $F_{ij}$ as the expected first time that a contact occurs between individuals $i$ and $j$ during a given night, given that a contact occurs. From Section 3.2.1 and , we can calculate an analytical value for $F_{ij}$ as: $F_{ij} = \sum_{t=0}^{t_{max}} tF_t(\mathbf{z}_{ij})$, where $t_{max} = 12$ hours and $F_t(\mathbf{z}_{ij})$ is the probability of the random walk, starting at the origin, being at the point $\mathbf{z}_{ij}$ for the first time after $t$ steps. As shown in Eq. (3.2), an expression for $F_t(\mathbf{z}_{ij})$ relies on $P_t(\mathbf{z}_{ij})$. However, as discussed when examining $P_{ij}$, the integral in Eq. (3.1) has so far proven intractable to evaluate. Therefore, we instead use random walk realisations to produce values of $F_{ij}$.

We perform the same random walk realisations as with the probability $P_{ij}$ to obtain a value for $F_{ij}$, but instead calculate the average first contact time for realisations where contacts occur. Figure 3.2B shows values of $F_{ij}$ for the integer distances and diffusion coefficients given in Section 3.3.1. The value of $F_{ij}$ increases as the distance between the centres of the individuals' home-ranges increases. For any given initial distance, as the diffusion coefficient increases, $F_{ij}$ decreases. Again, $\hat{d}_{ij}$ may not be an integer multiple of $\delta$ so, as with $P_{ij}$, we interpolate using the results from random walk realisations to obtain values of $F_{ij}$.

**Expected total time $T_{ij}$ spent in contact.**

We define $T_{ij}$ as the expected total time individuals $i$ and $j$ spend in contact during a given night. We calculate $T_{ij}$ by integrating the probability that a contact occurs at time $t$, for the time period between 0 and $t_{max} = 12$ hours. This probability is equivalent to the probability that the Euclidean distance between $i$ and

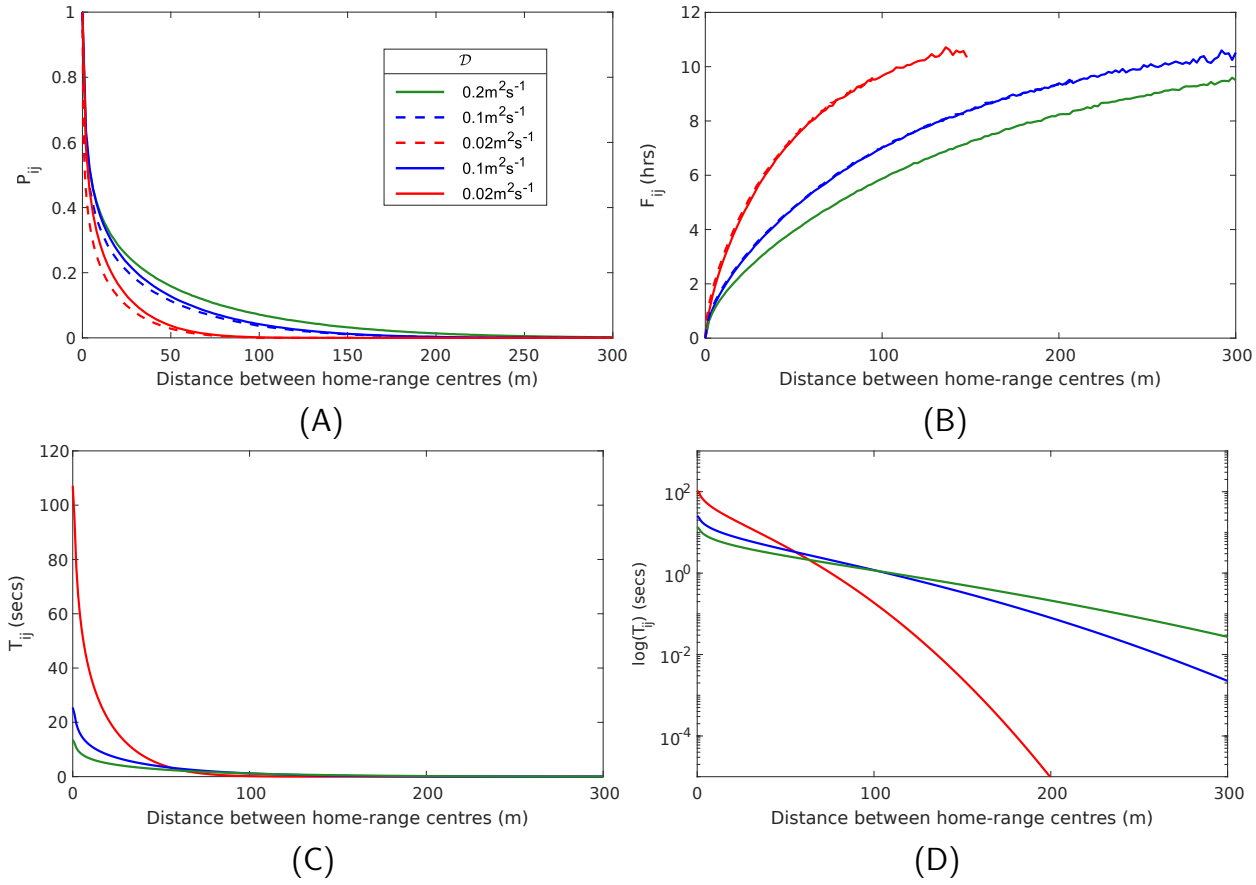



Figure 3.2: **For distances between 0 m and 300 m, and for diffusion coefficients 0.02m$^2$s$^{-1}$ (blue), 0.1m$^2$s$^{-1}$ (red), and 0.2m$^2$s$^{-1}$ (green), values of the nightly contact probability $P_{ij}$ (Fig. 3.2A), the expected first contact time $F_{ij}$ (Fig. 3.2B), and the total time spent in contact $T_{ij}$ (Figs. 3.2C and 3.2D) are calculated.** The dashed lines show the values of the contact measures $P_{ij}$ and $F_{ij}$ calculated by scaling the distances by $\sqrt{0.2/0.1}$ and $\sqrt{0.2/0.02}$, for 0.1 m$^2$s$^{-1}$ and 0.02 m$^2$s$^{-1}$ respectively. As the distance between the centres of the individuals' home-ranges increases, the values of both $P_{ij}$ and $T_{ij}$ decrease, whereas the value of $F_{ij}$ increases. Conversely, for any given initial distance, as the diffusion coefficient $\mathcal{D}$ increases, $F_{ij}$ decreases, whereas $P_{ij}$ increases. For small initial given distance, the value of $T_{ij}$ decreases as $\mathcal{D}$ increases, but as the initial distance increases, this relationship is inverted.

$j$ at time $t$, $\Delta_t$, is less than $\delta$, that is:

$$T_{ij} = \int_0^{t_{max}} P\left(\Delta_t < \delta\right)\ dt,$$

where $\Delta_0 = \hat{d}_{ij}$ is the distance between the home-range centres of $i$ and $j$. Defining $D_t = \Delta_t/\sqrt{2\mathcal{D}t}$, we can then write $T_{ij}$ as:

$$T_{ij} = \int_0^{t_{max}} P\left(D_t^2 < \frac{\delta^2}{2\mathcal{D}t}\right)\ dt.$$

From Section 3.2.1, we have that $D_t^2 \sim \chi^2(2, \hat{d}_{ij}^2/2\mathcal{D}t)$. Fig. 3.2C shows values of $T_{ij}$ for the range of distances (0 m to 500 m) and diffusion coefficients given in Section 3.3.1. As the distance $\hat{d}_{ij}$ between the centres of the individuals' home-ranges increases, the value of $T_{ij}$ decreases.

**Empirical values from data**

As detailed in Section 2.2.3, we can calculate analogous values of the three contact measures from empirical data, collected on populations of common brushtail possums. These possums populations lived in and around four non-contiguous study sites, designated A, B, C, and D, in mixed beech-podocarp forest in the Orongorongo Valley on the North Island of New Zealand (41°21′S 174°58′E, Fig. 2.1). During the field season, which ran from April to November 2012, trapping sessions occurred at regular intervals. Each trapping session involved traps being set and physical data of trapped possums collected. During the first trapping sessions, approximately 40 adult possums on each site were also collared with proximity loggers. These collars were designed to record any collar within a 1 metre distance.

Each record in this dataset corresponds to a single contact between a pair of possums. From these data, we can calculate the analogous value of the three contact measures. The nightly contact probability $P_{ij}$ is calculated as the as the proportion of nights, between the first and last contact (inclusive), during which a contact occurred. The average nightly first contact time $F_{ij}$ is calculated as the first time a record began recording, averaged over nights during which contacts occurred. The average total nocturnal time spent in contact $T_{ij}$ is calculated as the total nocturnal time spent in contact, averaged over nights during which contacts occurred.

We plot the measures against one another, and also plot the empirical values produced from the data. When examining these plots, shown in Fig. 3.3, we conclude that whilst there is a deterministic relationship between

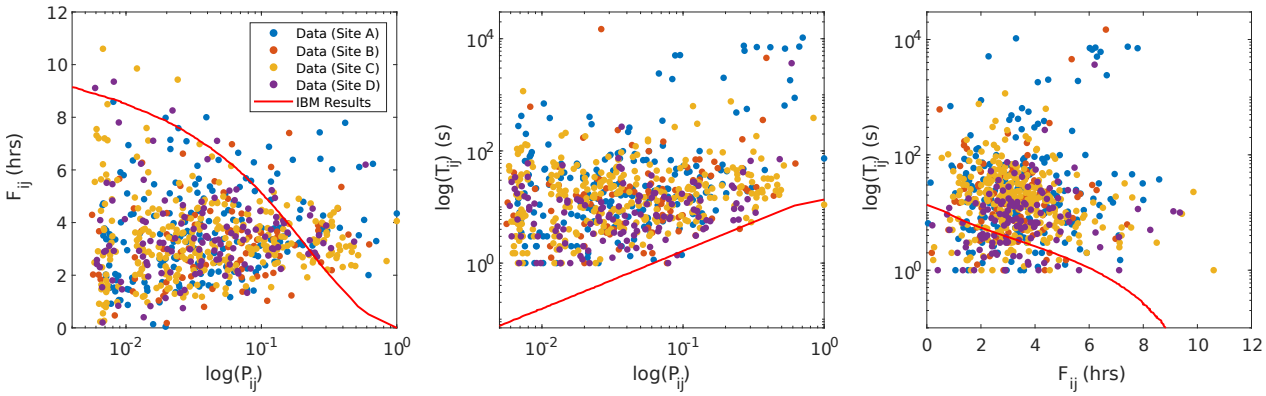



Figure 3.3: **Comparison of contact importance measures from model results and from empirical data.** From our model, and from the empirical data, we produce values of the three contact importance measures, the nightly contact probability $P_{ij}$, expected first contact time $F_{ij}$, and expected total time spent in contact $T_{ij}$. The coloured points represent the empirical data values, and the red line the model results. We conclude that the contact importance measures are deterministically related in the model but weakly correlated, if at all, in the empirical data. The exception is $P_{ij}$ and $T_{ij}$, which appear to be correlated in the empirical data.

$P_{ij}$, $F_{ij}$ and $T_{ij}$ (most likely arising because each measure is a function of the distance between home-range centres), there is at most a weak correlation between any pair of measures in the empirical data, with the exception of $P_{ij}$ and $T_{ij}$. This could indicate that our model is not adequately reflecting some aspect of possum movement or behaviour. For example, the lack of correlation between $P_{ij}$ and $F_{ij}$ could be explained by the topography of the sites. If there was an obstacle between the denning sites of two possums that lived close to one another, then they might not make contact very often, resulting in a low contact probability but, due to being close by, when they do come into contact the first contact time is also low. Furthermore, all three measures produced from our model exhibit a 'memory-less' property, as the movements on previous nights do not influence the movements during the current night. This is unrealistic, as it is likely that a possum will have preferred locations that they travel to in search of resources.

### 3.3.2 Effect of model assumptions on the contact measures

As discussed in the last part of Section 3.3.1, the discrepancy between the empirical values from the data, and those we produce from our model could indicate that our model is not adequately reflecting some aspect of possum movement or behaviour. The individual-based model presented in this section makes a variety of assumptions regarding the movements and behaviours of possums. Here, we examine these assumptions and consider the affect they exhibit on the values of the three contact measures previously defined and discussed.

#### No attraction or repulsion between possums

Possums are solitary animals, spending approximately 1% of their time interacting with other possums [Day et al., 2000]. Our model assumes that the movements of the possums are independent, that they do not change their behaviour regardless of any other possums in the vicinity. Introducing to the model some aspect of attraction between individuals would result in the values of the contact probability $P_{ij}$ and nightly contact time $T_{ij}$ increasing, and the value of the first contact time $F_{ij}$ decreasing, as possums make contact more often, stay in contact longer, and meet sooner during any given night. The opposite effect would be seen if some aspect of repulsion were introduced instead.

#### Lack of modelling behaviour when contacts occur

When two individuals in our model make contact, the behaviour during the contact is not modelled, and their subsequent movements are unaffected by the contact. The first two contact measures, $P_{ij}$ and $F_{ij}$, only consider the time from when a pair of individuals start moving to when they first come into contact with each other, if that happens at all. Any subsequent movements or behaviour are ignored, and do not affect their values. Conversely, the third measure $T_{ij}$ takes into account the entirety of the individuals' movements over a given night. This could therefore explain the lack of correlation between $P_{ij}$ and $T_{ij}$, and between $F_{ij}$ and $T_{ij}$, in addition to the greater values of $T_{ij}$ seen in the data compared with the model, as shown in Fig. 3.3.

If we were to model more realistically the behaviour exhibited during and after a contact, we would expect that $P_{ij}$ and $F_{ij}$ would remain unchanged, but $T_{ij}$ would be different. Whilst modelling their interactions could possibly increase the value of $T_{ij}$, modelling the behaviour subsequent to a contact could either increase or decrease $T_{ij}$, dependent on whether or not the two individuals would avoid each other afterwards.

### A nocturnal period lasting 12 hours

We have made the assumption that each individual spends a period of exactly 12 hours each night outside of their dens. This assumption, as mentioned in Section 3.3, is unrealistic. Whilst we will be using this assumption in our model, it is of worth to consider the effect that changing this 12 hour value will have on the contact measures we produce.

As the amount of time spent outside of dens increases, the values of the nightly contact probability $P_{ij}$, the expected first contact time $F_{ij}$, and the total time spent in contact $T_{ij}$ increase. For $P_{ij}$ and $T_{ij}$, this is most likely caused by the increase in the possibility of contacts occurring due to the additional time spent moving in the environment. As for $F_{ij}$, this is most likely caused by the increase in realisations during which contacts occur, allowing for individuals to meet later than previously allowed for in the model. Values of $P_{ij}$ (Fig. 3.4A), $F_{ij}$ (Fig. 3.4B), and $T_{ij}$ (Figs. 3.4C and 3.4D), are shown in Fig. 3.4. These values were calculated for a diffusion coefficient of $\mathcal{D} = 0.1$ m$^2$s$^{-1}$, a range of initial distances between individuals' home-range centres, and a range of times spent by the individuals outside their dens.

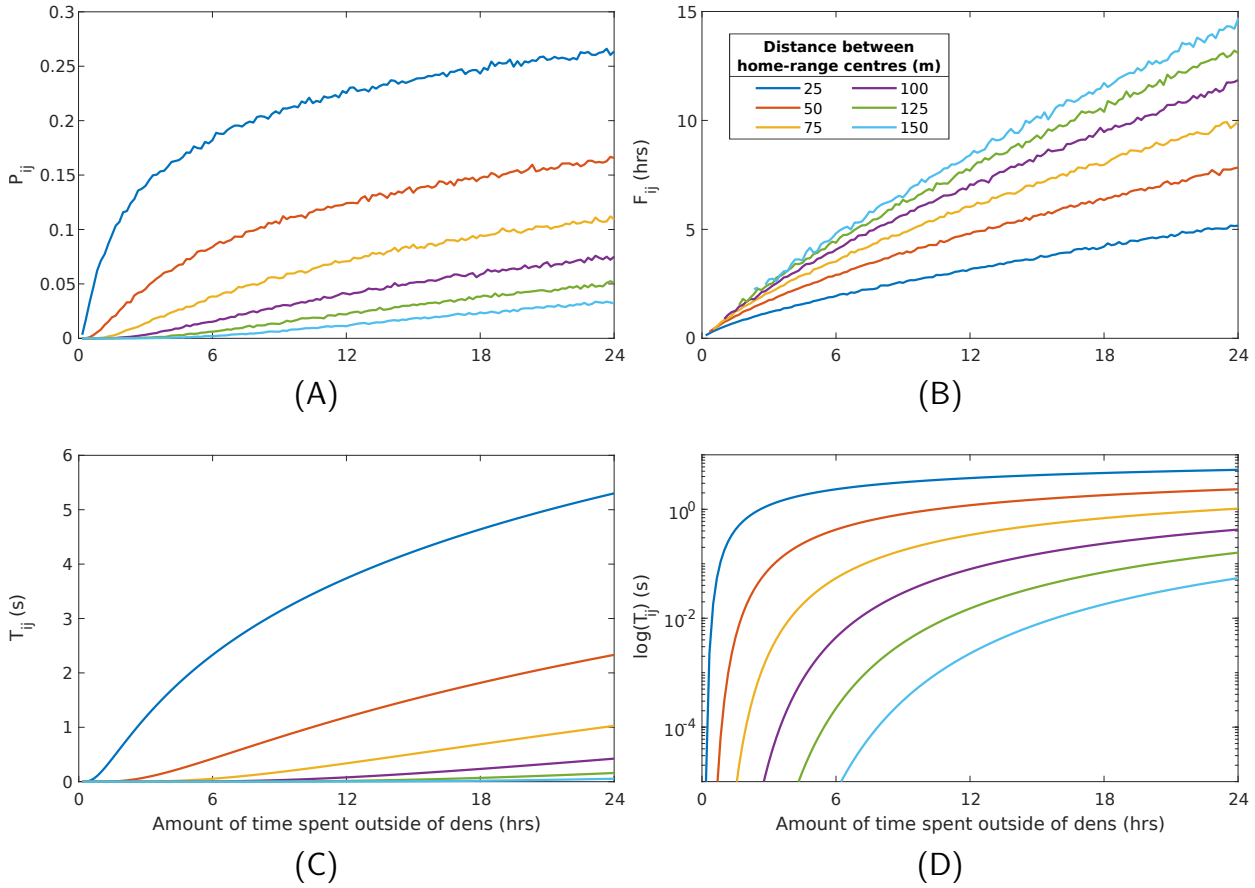



Figure 3.4: **As the amount of time spent outside of dens increases, the values of the nightly contact probability $P_{ij}$ (Fig. 3.4A), the expected first contact time $F_{ij}$ (Fig. 3.4B), and the total time spent in contact $T_{ij}$ (Figs. 3.4C and 3.4D) increase.** Across a range of initial distances between individuals' home-range centres, the values of $P_{ij}$, $F_{ij}$, and $T_{ij}$ are calculated for a range of times spent by the individuals outside their dens. The values were calculated for a diffusion coefficient of $\mathcal{D} = 0.1$ m$^2$s$^{-1}$. As the amount of time outside their dens increases, the values of $P_{ij}$ and $T_{ij}$ increase, due to the increased possibility of contacts occurring. Surprisingly, the value of $F_{ij}$ also increases. This is most likely correlated with the increase in $P_{ij}$, caused by the increase in realisations during which contacts occur.

### 3.3.3 Summary

In this section we have produced an individual-based model of possum movement and contact. Individuals move in their environment for 12 hours, according to an unbiased random walk on a lattice, starting from the centre of their circular home-range. From this model, we produce three measures of the length and importance of their contacts: the nightly contact probability $P_{ij}$, the average nightly first contact time $F_{ij}$, and the total nocturnal time spent in contact $T_{ij}$. Each of these are dependent on the distance between individuals' home-range centres, and the diffusion coefficient parameter. We go on to discuss methods, both analytical and, in the case of $P_{ij}$ and $F_{ij}$, simulated, for calculating these contact measures. Finally, after comparing with empirical values of the contact measures, we discuss ways in which our model assumptions could impact the model outputs.

## 3.4 Models of possum home-range centre distribution and collaring

In the model of possum movement described in Section 3.3, each individual in a population is assumed to have an identically sized, circular home-range. The model produces measures of contact length and importance that are functions of the distance between the centres of these home-ranges. In order to model a population of possums, a method for describing the distribution of their home-range centres in an environment is required.

Furthermore, the possum populations, whose contact data we use, are in fact a collared subset of larger populations consisting of trapped possums. As such, in additional to describing a model for the distribution of the home-range centres of a population of trapped individuals, in this section we go on to model the selection of a collared subset.

### 3.4.1 Home-range centre distribution model

To model a population of trapped individuals, $N$ points, representing the individuals' home-range centres, are placed in a circle of radius $R_I$, centred at the origin. We follow the methodology of Richardson et al. (2017), distributing the home-range centres using a homogeneous spatial Poisson point process with density parameter $\lambda > 0$. Here $R_I = \sqrt{N/\pi\lambda}$. For each site, values of the parameters used in these models ($N$, $C$, $R_I$ and $\lambda$) were taken from data (see Table 3.1).

The possum pair contact data, described in Chapter 2, has no accompanying information about the locations of the possums' dens, nor of the topography of the areas they inhabited. Because of this, a homogeneous spatial Poisson point process is the most appropriate model to use. As discussed in Efford and Fewster (2013), spatial Poisson point processes, both homogeneous and inhomogeneous, have been used to distribute home-range centres, due to computational convenience. However, other methods have been used. In Reich and Gardner (2014), a Strauss process, which involves repulsion between individuals' home-ranges, is used to model the home-range centre distribution.

### 3.4.2 Models of possum collaring

Here we describe three different models for collaring a subset of a population of trapped individuals. Each model assigns to an individual a probability of being collared, and a subset of the trapped individuals are sampled using this collaring probability. During the trapping sessions of the possums in the data, which

amounted to $k = 16$ nights, $N$ possums were trapped by any of the $n_{\text{traps}} = 100$ traps arranged in a square grid, of spacing 40 m and covering  13 ha, and $C$ of those possums were then collared. To each individual $i$ we assign a collaring probability $c_i$, calculated using a trapping function evaluated at $\mathbf{z}_i$, the centre of individual $i$'s circular home-range. We collar $C$ individuals by sampling individuals without replacement, using the values of $c_i$ as weights.

### Model 1: Uniform collaring probability

We assume all individuals in a population are equally likely to be collared. This gives us a uniform collaring probability, where for a given individual $i$, $c_i = \dfrac{1}{N}$. For clarity, we refer to this model as the Uniform Collaring Model.

### Model 2: Central trapping function $T_{\text{central}}$

For clarity, we refer to this model as the Central Trapping Model. We assume that an individual's likelihood of being collared is determined by the probability of them being trapped at any point during an observation period. An individual whose home-range centre is at the origin is expected to be trapped on each of the $k$ nights of the trapping sessions whereas, if the home-range centre is Euclidean distance $R_I$ from the origin, the individual is expected to be trapped on only one of the $k$ nights. We also assume that the probability of an individual being trapped is an exponential function of their distance from the origin. The collaring probability $c_i$ is therefore equal to this central trapping function $T_{\text{central}}$, evaluated at $\mathbf{z}_i$. Hence, for a given individual $i$,

$$c_i = T_{\text{central}}(\mathbf{z}_i) = k^{\left(-\frac{\|\mathbf{z}_i\|_2}{R_I}\right)^2}, \text{ where } k = 16.$$

Fig. 3.5A shows the probability density function of $T_{\text{central}}$ for $R_I = 300$ m.

### Model 3: Grid trapping function $T_{\text{grid}}$

For future clarity, we refer to this model as the Grid Trapping Model. As with the Central Trapping Model, it is assumed that the collaring probability, $c_i$, is determined by the probability of them being trapped at any point during an observation period. We assume individuals whose home-range centres are closer to the grid of $n_{\text{traps}}$ traps are trapped more frequently that those further away.

In Efford (2004), a two-parameter model for the detection probability of a single animal by a single trap is

described, as a half-normal function of the distance between the animal's home-range centre and the trap's location. The model has parameters $g_0$, the detection probability at the home-range centre, and $\sigma$, which describes the spatial scale over which the detection or trapping probability declines. For an individual $i$, with home-range centre $\mathbf{z}_i$, and a trap at location $\mathbf{y}$, the detection probability is given by:

$$P_{detection}(\mathbf{z}_i, \mathbf{y}) = g_0 \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{y}\|_2^2}{2\sigma^2}\right).$$

We build on this model by describing a function $T_{\text{grid}}$, that gives the probability that individual $i$ is trapped at least once by any of the grid of $n_{\text{traps}}$ traps at least once during the $k = 16$ nights of the trapping sessions. We arrange the $n_{\text{traps}}$ traps on a square lattice, with $\mathbf{y}_j$ denoting the location of the $j^{th}$ trap. The collaring probability $c_i$ is equal to this grid trapping function $T_{\text{grid}}$, evaluated at $\mathbf{z}_i$, that is:

$$c_i = T_{\text{grid}}(\mathbf{z}_i) = 1 - \left[\prod_{j=1}^{n_{\text{traps}}} \left(1 - P_{detection}(\mathbf{z}_i, \mathbf{y}_j)\right)\right]^k.$$

Values of $g_0$ and $\sigma$ for the specific possum populations we are studying are estimated in Richardson et al. (2017), and are shown in Table 3.2. The traps are laid on a square lattice of length 360 metres, at a granularity of 40 metres. This distribution of traps was also estimated from Richardson et al. (2017). Fig. 3.5B shows the probability density function of $T_{\text{central}}$ for $R_I = 300$ m, $k = 16$, $g_0 = 0.08$ and $\sigma = 48m$.

### 3.4.3 Summary

We have described a model for the distribution of home-range centres of a population of $N$ trapped individuals. We have also discussed three different models for producing a population of $C$ collared individuals from these trapped individuals. Each collaring model attempts to recreate the distribution of the home-range centres of the collared possums for whom we have contact data. Example populations of 40 individuals produced using the three collaring models are shown in Fig. 3.6. For each, an inhabiting radius $R_I$ of 300m was used. In the Grid Trapping Model, values of $g_0 = 0.08$ and $\sigma = 48m$ are used.

The locations of the individuals' home-range centers are more concentrated in the Central Trapping Model than the Grid Trapping Model. This is a result of the trapping functions used, with an illustration given in Fig. 3.5. The distribution of possum home-range centres differs little between the Uniform Collaring Model and the Grid Trapping Model. This is due to the value of the grid trapping function $T_{\text{grid}}$ being close to 1

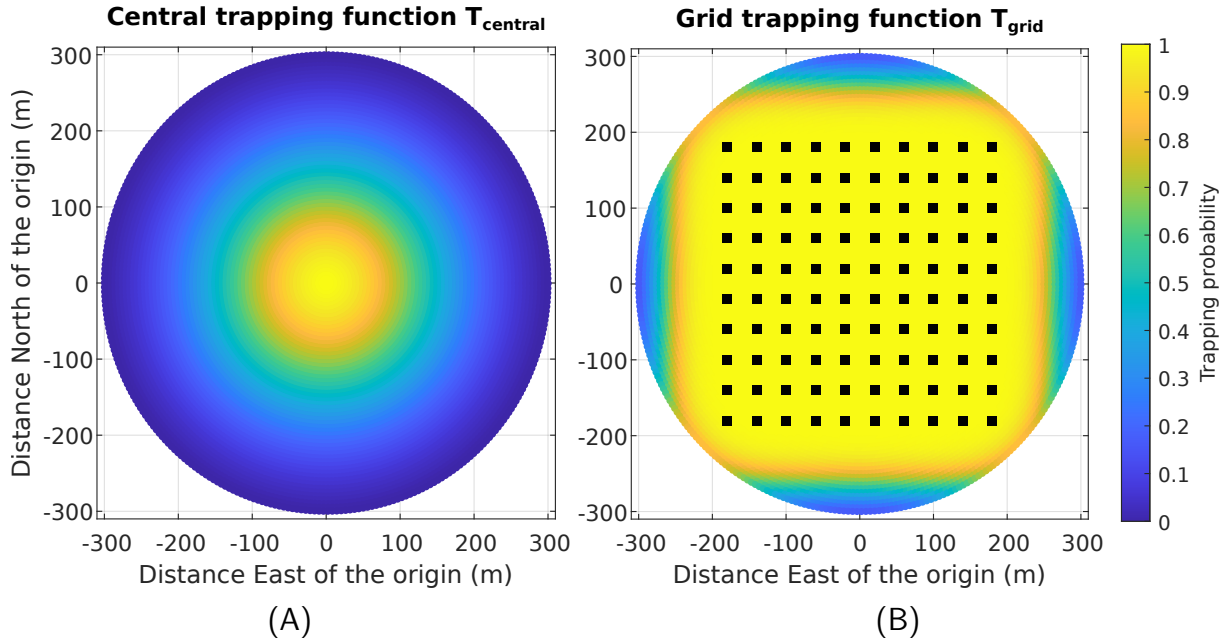



Figure 3.5: **When the central trapping function $T_{\text{central}}$ is used to collar individuals, their locations are more tightly centred about the origin than when the grid trapping function $T_{\text{grid}}$ is used.** The trapping functions, $T_{\text{central}}$ and $T_{\text{grid}}$, are shown above. For both, an inhabiting radius $R_I$ of 300m and $k = 16$ are used. For $T_{\text{grid}}$, values of $g_0 = 0.08$ and $\sigma = 48m$. Further, the locations of the traps used are shown as black squares. Because of the wide distribution of the traps in the individual's inhabiting area, the probability of being trapped is higher throughout the inhabiting area than when the $T_{\text{central}}$ is used.

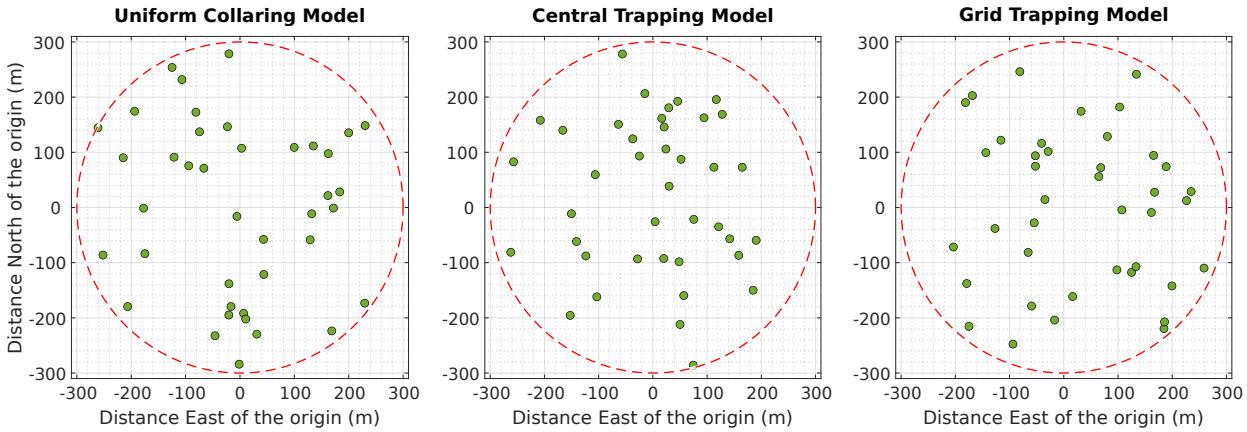



Figure 3.6: **Example populations of 40 individuals produced using the three collaring models.** For each, an inhabiting radius $R_I$ of 300m was used. In the Grid Trapping Model, values of $g_0 = 0.08$ and $\sigma = 48m$ are used. The locations of the individuals' home-range centers are more concentrated in the Central Trapping Model than in either the Grid Trapping Model or the Uniform Collaring Model. This is a result of the trapping functions used, with an illustration given in Fig. 3.5.

almost everywhere except for the boundaries of the study area. The goal of this is to provide a mechanism for linking the contact results to the distribution of possums in an environment, something that we are unable to do using the empirical data alone. Later in this chapter, we will build contact networks (both binary and weighted) to compare with empirical networks.

## 3.5 Estimating diffusion coefficients and home-range size from data

The diffusion coefficient $\mathcal{D}$ is a necessary parameter for the individual based model discussed in Section 3.3. We produce distributions of $\mathcal{D}$ for each of the four study sites (which we examine separately), that will allow our model to match closely with the data in Section 2.2. This is achieved by fitting pairwise distances between collared individuals in a model population to estimated distances between possum pairs in the data. We employ a procedure to estimate a single optimal value of the diffusion coefficient $\mathcal{D}$. To produce distributions of $\mathcal{D}$, we perform this procedure many times, each with a different model population.

### 3.5.1 Procedure for estimating a diffusion coefficient

The following four steps are an outline of the procedure for estimating a diffusion coefficient. Each step will subsequently be described in detail (see Fig. 3.9 for visual representation of this procedure).

1. Simulate the spatial locations of a model population's home-range centres, and the subset of individuals that are collared. Produce the set $d_{model}$ of pairwise distances between these model collared individuals.

2. For a given value of $\mathcal{D}$, use contact probabilities of possum pairs in the data to produce a set $d_{data}$ of estimated distances between their home-range centres. The contact probabilities and their associated distances are dependent on $\mathcal{D}$.

3. Quantify the difference between the distributions of $d_{model}$ and $d_{data}$.

4. Repeat Steps 2 and 3 for each value of $\mathcal{D}$ in a predetermined set of discrete values ranging from 0 $\text{m}^2\text{s}^{-1}$ to 0.5 $\text{m}^2\text{s}^{-1}$. Determine the optimal value of $\mathcal{D}$ such that the difference between the empirical cumulative distribution functions of $d_{model}$ and $d_{data}$ is minimised.

We repeat Steps 1-4 a set number of times to produce the distribution of $\mathcal{D}$.

**Step 1: Produce a set of pairwise distances, $d_{model}$, between the simulated home-range centres of a model population of $C$ collared individuals.**

We simulate the home-range centre locations of $C$ collared individuals, sampled from a larger population of $N$ trapped individuals distributed as per Section 3.4.1. The parameters used are given in Table 3.1. Calculating the pairwise distances between these home-range centres produces a set of distances $d_{model} = \{d_{ij} \mid i, j \text{ individuals in the model population}\}$.

| Site | N° possums trapped $N$ | N° possums collared $C$ | Estimated density $\lambda$ ($m^{-2}$) | Inhabiting area radius $R_I$ (m) | N° possum pairs who made contact | N° possible pairs | Scaling factor $\omega$ |
|---|---|---|---|---|---|---|---|
| A | 146 | 39 | $4.87\times10^{-4}$ | 308.91 | 189 | 741 | 0.2551 |
| B | 178 | 41 | $6.92\times10^{-4}$ | 286.14 | 88 | 820 | 0.1073 |
| C | 144 | 40 | $4.08\times10^{-4}$ | 335.18 | 293 | 780 | 0.3756 |
| D | 140 | 39 | $4.20\times10^{-4}$ | 325.74 | 128 | 666 | 0.1922 |

Table 3.1: **Summary data of the possum populations**. Values, pertaining to the real-world possum populations discussed in Chapter 2. The values of $N$, $C$, $\lambda$ are taken directly from Richardson et al. (2017), whereas $R_I$ is calculated as $R_I = \sqrt{N/\pi\lambda}$). The number of pairs who made contact is drawn from Table 2.5. The number of possible pairs is calculated as $\frac{C(C-1)}{2}$. The exception is site D. There were two possums who had no contact with any other possum during the field study. Therefore, we use a value of $C = 37$ when calculating the number of possible pairs. The scaling factor $\omega$ is calculated as the quotient of the preceding two columns.

**Step 2: Estimating pairwise distances $d_{data}$ between possums in the data for a given value of $\mathcal{D}$**

For possum pairs in the data, we have estimated values of the contact probability $P_{ij}$ as the ratio of nights where contacts occurred to the total nights that the possum pairs could have met. From these contact probabilities we use the relationship shown in Fig. 3.2A to estimate the Euclidean distance between pair's home-range centres from the contact probabilities, for a given value of $\mathcal{D}$. This gives us a set of distances, $d_{data} = \{d_{ij} \mid i, j \text{ possums in the data who made contact}\}$, one for each possum pair that had a contact.

**Step 3: Quantifying the difference between the distributions of $d_{model}$ and $d_{data}$**

To quantify the difference between the distributions of pairwise distances from our model population and from the real-world possums, we calculate the mean squared error (MSE) between their empirical cumulative distribution functions (eCDFs).

Firstly, we produce the eCDFs of both $d_{model}$ and $d_{data}$. We refer to them respectively as $f_{model}$ and $f_{data}$. These functions are evaluated at the unique values of $d_{model}$ and $d_{data}$ respectively, which we denote as $U_{model}$ and $U_{data}$.

The set of distances produced in Step 1 represents the entirety of all pairs in a model population, whereas the set of distances produced in Step 2 represents only a proportion of all possible real-world possum pairs, as only some pairs made contact during the study period. When comparing $f_{model}$ and $f_{data}$, it is therefore necessary to scale $f_{data}$ by the proportion, $\omega$, of possum pairs it represents. The values of $\omega$ are given in Table 3.1, and calculated as the quotient of the number of pairs who made contact (drawn from Table 2.5),

and the number of possible pairs. The number of possible possum pairs is calculated as $\frac{C(C-1)}{2}$. A small caveat is site D, where there were two possums who had no contact with any other possum during the field study. Therefore, we use a value of $C = 37$ when calculating the number of possible pairs.

To calculate the MSE between the eCDFs $f_{model}$ and $f_{data}$, we first define the error between $f_{model}$ and $f_{data}$ for a given pairwise distance $d_{ij}$, calculated as

$$\text{Error}(d_{ij}) = \|\omega f_{data}(d_{ij}) - f_{model}(d_{ij})\|_1 .$$

The MSE is the sum of the squares of these errors, evaluated at the unique values of $d_{data}$, $U_{data}$, and of $d_{model}$, $U_{model}$. The formula for the mean squared error between $d_{model}$ and $d_{data}$ is given in Eq. (3.5). We discard any errors calculated for $d_{ij} \in U_{model}$ where $d_{ij} > \max(U_{data})$, as the value of $\text{Error}(d_{ij})$ at these points is meaningless. Because of the meaninglessness of the discarded errors, we cannot use the two-sample Kolmogorov-Smirnov test statistic, as defined in Section 3.2.3. However, the maximal value of the non-discarded errors is analogous to this test statistic.

$$\text{MSE}(d_{model}, d_{data}) = \frac{1}{|U_{model}| + |U_{data}|} \left( \sum_{d_{ij} \in U_{data}} \left(\text{Error}\left(d_{ij}\right)\right)^2 + \sum_{\substack{d_{ij} \in U_{model} \\ d_{ij} \leq \max(U_{data})}} \left(\text{Error}\left(d_{ij}\right)\right)^2 \right) \tag{3.5}$$

The process of calculating the MSE is illustrated in Fig. 3.7, where the exemplar MSE equals $2.67 \times 10^{-3}$, and the maximal error equals $8.10 \times 10^{-2}$.

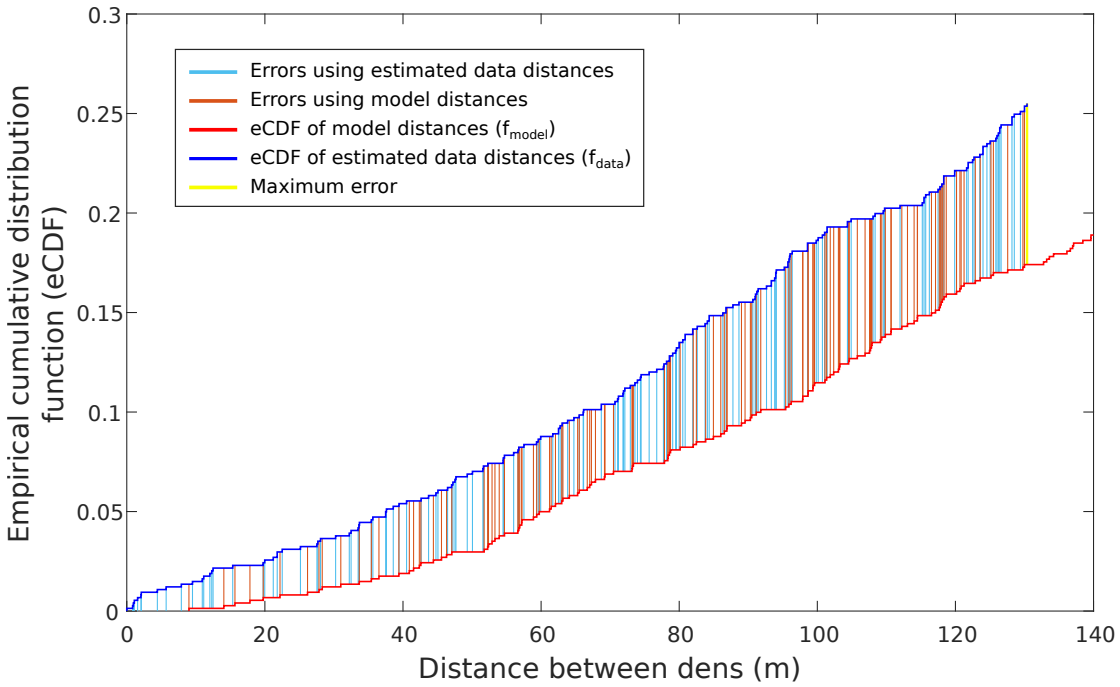



Figure 3.7: **An illustration of the method for quantifying the difference between the distributions of $d_{data}$ and $d_{model}$.** For $\mathcal{D} = 0.06\text{m}^2\text{s}^{-1}$, we calculate the pairwise distances $d_{data}$ between individuals in the data. Simultaneously, we create a single model population, using the Uniform Collaring Model and parameters for site A, and determine pairwise distances $d_{model}$. We produce the empirical cumulative distribution functions (eCDFs) of $d_{model}$ and $d_{data}$, respectively $f_{model}$ and $f_{data}$, and calculate the mean squared error between them. This is done by calculating the squares of the lengths of the light blue and orange lines, as detailed in Section 3.5.1. The maximal error is shown in yellow and is analogous to the two-sample Kolmogorov-Smirnov test statistic. In this example the MSE = $2.67\times10^{-3}$, and the maximal error equals $8.10\times10^{-2}$.

**Step 4: Finding global optimal value of $\mathcal{D}$.**

We perform Steps 2 and 3 on every value of the diffusion coefficient $\mathcal{D}$ in a predetermined set of values. When calculating the mean squared error for these $\mathcal{D}$, there may not exist a singular optimal value of the diffusion coefficient. Instead, there is a range of values in our predetermined set that produce the minimal mean squared error. This is because of the granularity of the diffusion coefficients. For values of $\mathcal{D}$ close to one another, the values of the error function are the same, meaning that the MSE overall will also be the same. In a handful of realisations, the minimal range is non-contiguous. Due to the infrequency of this occurring, if the minimal range is non-contiguous, we repeat Steps 1-3 until a contiguous minimal range is produced. Only then is the minimal range recorded and the repetition number advanced by 1. We take the mid-range value as our optimal diffusion coefficient $\mathcal{D}$.

Fig. 3.8 shows the mean-squared errors (in black) calculated for a single population, using the Uniform Collaring Model and parameters for site A, and for diffusion coefficients between 0 $m^2s^{-1}$ to 0.5 $m^2s^{-1}$. The mean-squared errors are shown in black. The minimum value of the MSE is $1.0379\times10^{-4}$, shown by the blue dashed line. The contiguous minimal range of $\mathcal{D}$ is [0.14189 $m^2s^{-1}$, 0.14192 $m^2s^{-1}$], with the optimal value of the diffusion coefficient being the mid-range value 0.141905 $m^2s^{-1}$, shown in red.

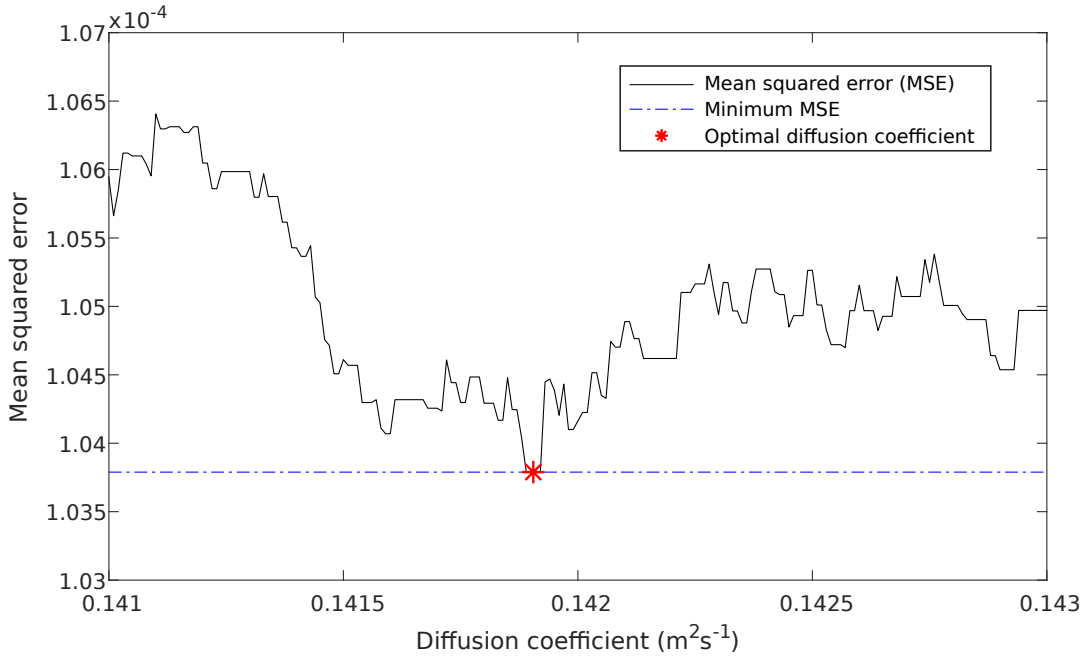



Figure 3.8: **There exists a contiguous global optimal range of values for the diffusion coefficient $\mathcal{D}$ that gives the minimum mean squared error.** We create a single population, using the Uniform Collaring Model and parameters for site A, and perform Step 3 for diffusion coefficients between 0 $m^2s^{-1}$ to 0.5 $m^2s^{-1}$. The mean-squared errors are shown in black. The minimum value of the MSE is $1.0379\times10^{-4}$, shown by the blue dashed line. The contiguous minimal range of $\mathcal{D}$ is [0.14189 $m^2s^{-1}$, 0.14192 $m^2s^{-1}$], with the optimal value of the diffusion coefficient being the mid-range value 0.141905 $m^2s^{-1}$.

**Producing a distribution of $\mathcal{D}$**

We repeat Steps 1-4 for a set number of times. Each repetition uses a different model population. These repetitions result in a distribution of diffusion coefficients $\mathcal{D}$ that allow our model to match closely with the data. This method is illustrated in the flowchart shown in Fig. 3.9.

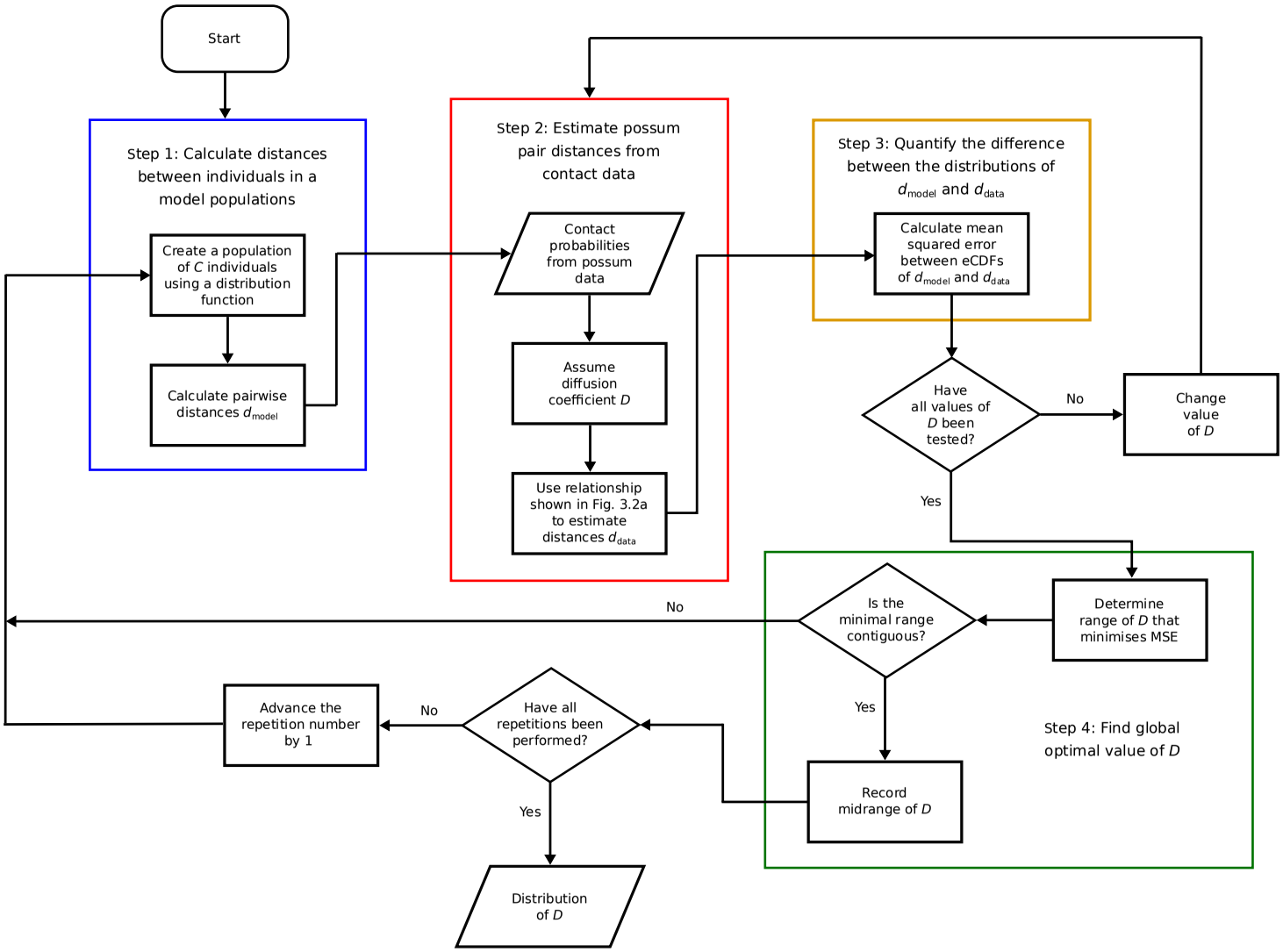



Figure 3.9: **Flowchart of the method, described in Section 3.5.1, of estimating $\mathcal{D}$ from the data described in Section 2.2.** For a predetermined set of values of $\mathcal{D}$, this method calculates the mean squared error (MSE) between the empirical cumulative distribution functions (eCDFs) of $d_{data}$ and $d_{model}$. We determine the range of $\mathcal{D}$ that minimises the MSE and, if the range is contiguous, record the midrange value as our optimal $\mathcal{D}$. The steps described in Section 3.5 are highlighted, Step 1 in blue, Step 2 in red, Step 3 in yellow, and Step 4 in green. This method produces a distribution of $\mathcal{D}$ that minimise allow our model to match closely with the data.

### 3.5.2 Diffusion coefficient estimation results

Fig. 3.10 shows four distributions of estimated diffusion coefficients for sites A to D, corresponding to the possum home-range distribution models. The variation in $\mathcal{D}$ is a result of the stochastic nature of the method employed to create a model population. Each distribution was produced using the method detailed in the previous section, with $10^4$ repetitions of Steps 1-4. Values of $\mathcal{D}$ between 0 $m^2s^{-1}$ and 0.5 $m^2s^{-1}$, with steps of $1 \times 10^{-5}$ $m^2s^{-1}$, were tested. Table 3.2 shows the mean ($\mu_{\mathcal{D}}$) for each of the diffusion coefficient distributions, with the standard deviation in brackets. Whilst the results for sites A and D are similar, there are clear differences between each site. As shown in Fig. 2.6A, the distributions of empirical values of $P_{ij}$ are similar enough across the four sites that they are unlikely to cause the differences in minimal diffusion coefficient distributions. Instead, the differences are most likely caused by the values of $N$, $C$, $\lambda$, $R_I$ and other parameters that are used to create the model populations.

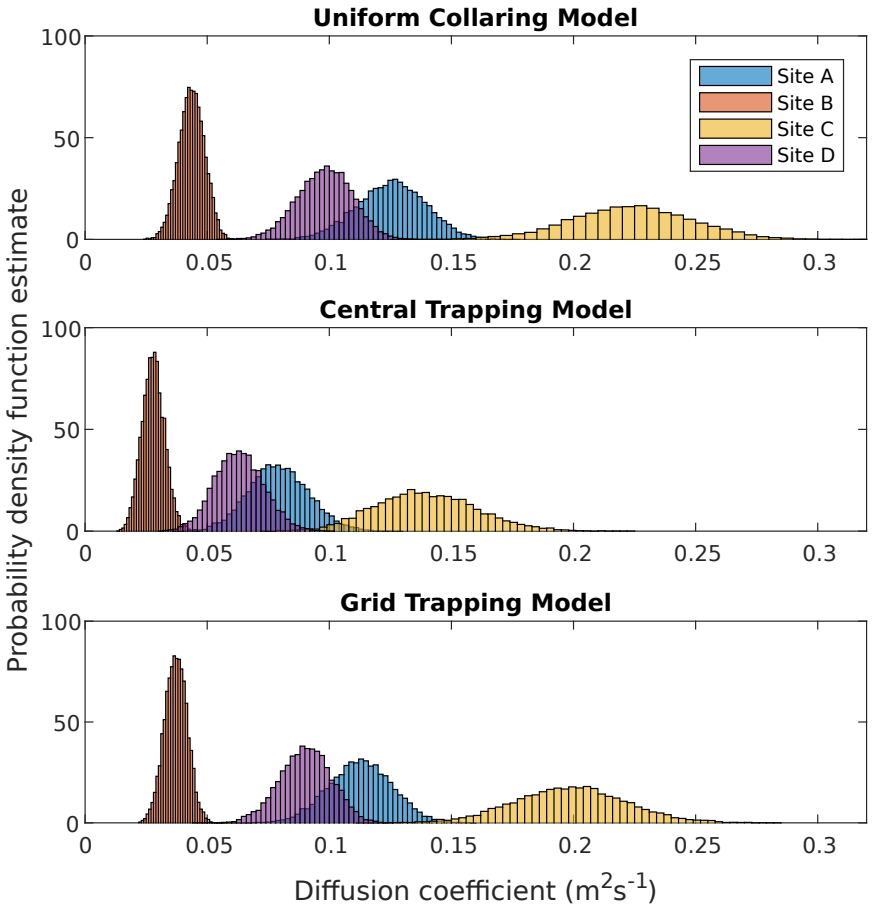



Figure 3.10: **Distributions of $\mathcal{D}$ estimated for each of the four possum distribution models.** The diffusion coefficient distributions are broadly similar between the distribution models. The values produced using the Uniform Collaring Model are consistently larger site to site when compared to any other model.

Local minima do occur, but in each repetition, a well-defined global minimal range was found. Therefore a grid-search, rather than an optimisation, approach was implemented. During the production of these

| | | | Mean optimal diffusion coefficient $\mu_{\mathcal{D}}$ ($m^2s^{-1}$) | | |
|---|---|---|---|---|---|
| Site | $g_0$ | $\sigma$ (m) | Uniform Collaring Model | Central Trapping Model | Grid Trapping Model |
| A | 0.088 | 45.82 (8.81) | 0.12542 (0.014098) | 0.07932 (0.012297) | 0.11274 (0.012899) |
| B | 0.105 | 31.68 (6.28) | 0.04361 (0.005276) | 0.02791 (0.004635) | 0.03717 (0.004644) |
| C | 0.054 | 60.03 (16.00) | 0.22366 (0.024693) | 0.13955 (0.020600) | 0.19869 (0.022797) |
| D | 0.059 | 59.29 (12.99) | 0.09819 (0.011129) | 0.06309 (0.009992) | 0.09023 (0.010482) |

Table 3.2: **The mean optimal value of the diffusion coefficient is smallest when using the Central Trapping Model, and greatest when using the Uniform Collaring Model.** Values of $g_0$, $\sigma$ for each site, obtained or calculated using values from Richardson et al. (2017). Further, the mean optimal diffusion coefficient is calculated (s.d. is given in brackets). The distribution of these diffusion coefficient is shown in Fig. 3.10.

distributions of $\mathcal{D}$, approximately 0.23% of repetitions performed resulted in a non-contiguous minimal range. As mentioned previously, when this occurred the procedure was repeated until a contiguous minimal range was produced. We chose not to use the maximal error to determine the optimal diffusion coefficient, as the minimal range of $\mathcal{D}$ determined by the maximal error is between 42 and 250 times as large (depending on the site and possum collaring model) as that determined using the mean squared error.

Fig. 3.11 shows the distributions of the minimal diffusion coefficient range sizes. The most common outcome of a repetition is that of a unique optimal $\mathcal{D}$, with the Central Trapping Model resulting in the highest proportion of realisations where this occurs. This is most likely due to the shape of the empirical cumulative distribution function of the model population, resulting in fewer values of $\mathcal{D}$ with the same values of the error function. The distributions from site C of minimal $\mathcal{D}$ range consistently have the highest proportion of realisations with unique $D$, with site B consistently having the lowest proportion. Conversely, the maximal value of the distributions from site B are larger than site A, C and D. This is most likely due to the shape of the eCDF of the estimated pairwise distances from the data.

The mean estimated diffusion coefficients produce sensible percentages of time spent by possums travelling. Using the parameter values, $\delta = 1$ metre and $\tau = 1$ second, we have that the diffusion coefficient is $\mathcal{D} = p/2$. This therefore means that the proportion of time steps where an individual moves from its position is $2\mathcal{D}$, which is analogous to the real-world nocturnal time spent travelling. Table 3.3 shows the percentages of time spent travelling, calculated using the mean values given in Table 3.2. Maclennan (1984) found that possums, living in an open eucalypt woodland near Lake Manchester in Australia, spent approximately 30% of the time outside their dens travelling. Whilst the possums we study live in a different environment to those near Lake

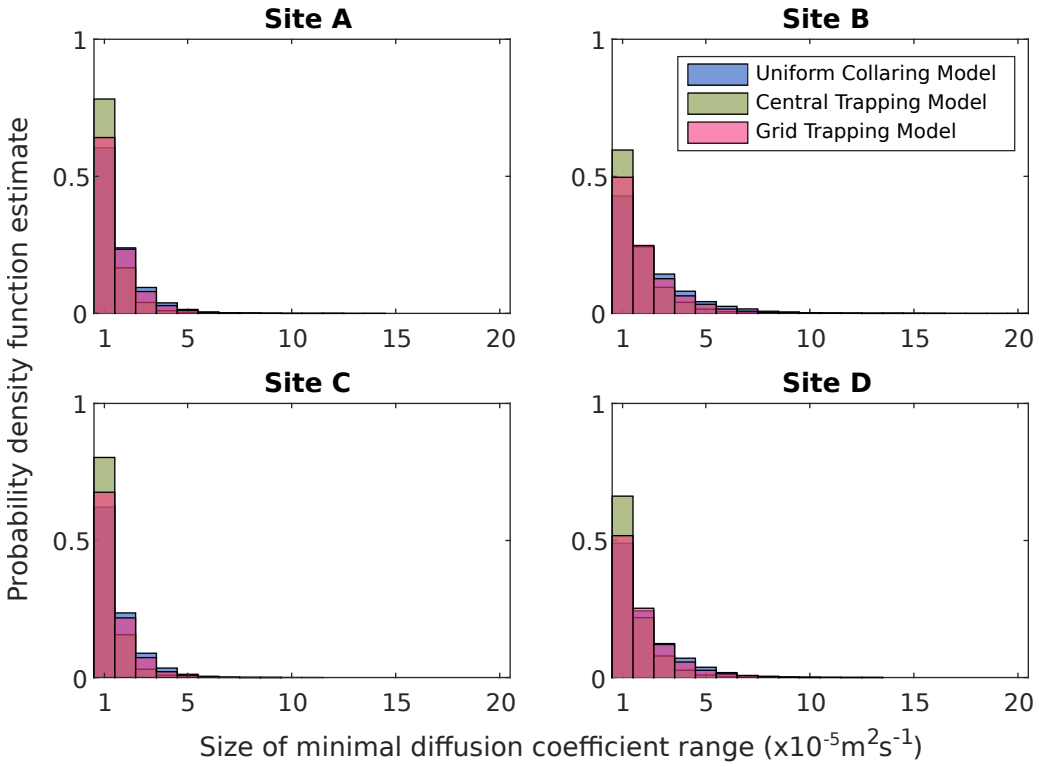



Figure 3.11: **The most common outcome of a realisation is a unique optimal diffusion coefficient $\mathcal{D}$.** Distributions of the minimal diffusion coefficient range sizes are shown. The Central Trapping Model results in the highest proportion of realisations where a unique $\mathcal{D}$ is produced. The distributions from site C of minimal $\mathcal{D}$ range consistently have the highest proportion of realisations with unique $D$, with site B consistently having the lowest proportion. Conversely, the maximal value of the distributions from site B are larger than site A, C and D.

Manchester, the percentages of time spent by possums travelling, calculated using our estimated diffusion coefficients, are similar. This gives some indication that our model replicates some aspects of real-world possum behaviour.

We compare the distributions of diffusion coefficients with those from the Uniform Collaring Model, which are consistently larger across each site. The values of $\mu_{\mathcal{D}}$ from the Uniform Collaring Model are on average 57.6% greater than the Central Trapping Model and 12.5% greater than the Grid Trapping Model. Values of the standard deviation from the Uniform Collaring Model are on average 14.9% greater than the Central Trapping Model and 9.3% greater than the Grid Trapping Model.

This difference can be explained by the distribution of individuals' home-range centres. Comparing to the Uniform Collaring Model, locations of the home-range centres in each of the other models are more concentrated about either the origin in the Grid Trapping Model and the Central Trapping Model. This decreases the values of the distances $d_{model}$ between pairs of collared individuals, which in turn means a

| | Percentage of nocturnal time spent travelling | | |
|---|---|---|---|
| Site | Uniform collaring probability | Central trapping | Grid trapping |
| A | 25.08% | 15.86% | 22.55% |
| B | 8.72% | 5.58% | 7.43% |
| C | 44.73% | 27.91% | 39.74% |
| D | 19.64% | 12.62% | 18.05% |

Table 3.3: **The mean estimated diffusion coefficients produce sensible percentages of time spent by possums travelling.** As $\mathcal{D} = p/2$, we have that the proportion of nocturnal time spent travelling is given by $2\mathcal{D}$. Maclennan (1984) found that possums, living in an open eucalypt woodland near Lake Manchester in Australia, spent approximately 30% of their time travelling. Comparing the proportions calculated using the mean values given in Table 3.2 with this real-world information shows our estimated diffusion coefficients give sensible percentages of time spent by possums travelling. It should be noted that the possums we study live in different conditions to those in Australia.

smaller value of $\mathcal{D}$ is required in order for the MSE between the CDFs of $d_{model}$ and $d_{data}$ to be minimised. Distributions of the central and grid trapping functions are illustrated in Fig. 3.5. The comparatively small difference between the Uniform Collaring Model and the Grid Trapping Model is due in part to the trapping function being close to 1 at almost all points in the distribution area. This in turn, causes only a slight difference between the distributions of home-range centres between the Grid Trapping Model and the Uniform Collaring Model.

Using the values of $\mu_{\mathcal{D}}$ from each model, given in Table 3.2, we produce sets of pairwise distances for idealised model populations of size $C \times 10^4$ and calculate the mean squared error between the eCDFs of these sets and of the data. The mean squared errors are given in Table 3.4, and the eCDFs are shown in Fig. 3.12. The Uniform Collaring Model gives the smallest MSE for all sites, whilst the Central Trapping Model gives the largest MSE.

| | Mean squared error | | |
|---|---|---|---|
| Site | Uniform collaring probability | Central trapping | Grid trapping |
| A | $2.708\times10^{-4}$ | $3.474\times10^{-4}$ | $2.737\times10^{-4}$ |
| B | $1.888\times10^{-4}$ | $2.175\times10^{-4}$ | $1.925\times10^{-4}$ |
| C | $6.904\times10^{-4}$ | $8.389\times10^{-4}$ | $7.344\times10^{-4}$ |
| D | $1.900\times10^{-4}$ | $2.485\times10^{-4}$ | $1.981\times10^{-4}$ |

Table 3.4: **The mean squared error between the empirical cumulative distribution functions of pairwise distances between individuals in an idealised model populations of size $C \times 10^4$, and of estimated distances between possum pairs in the data produced using the mean optimal value of the diffusion coefficient given in Table 3.2.** The value of the mean squared error is greatest when using the Central Trapping Model, and smallest when using the Uniform Collaring Model. The eCDFs are shown in Fig. 3.12.

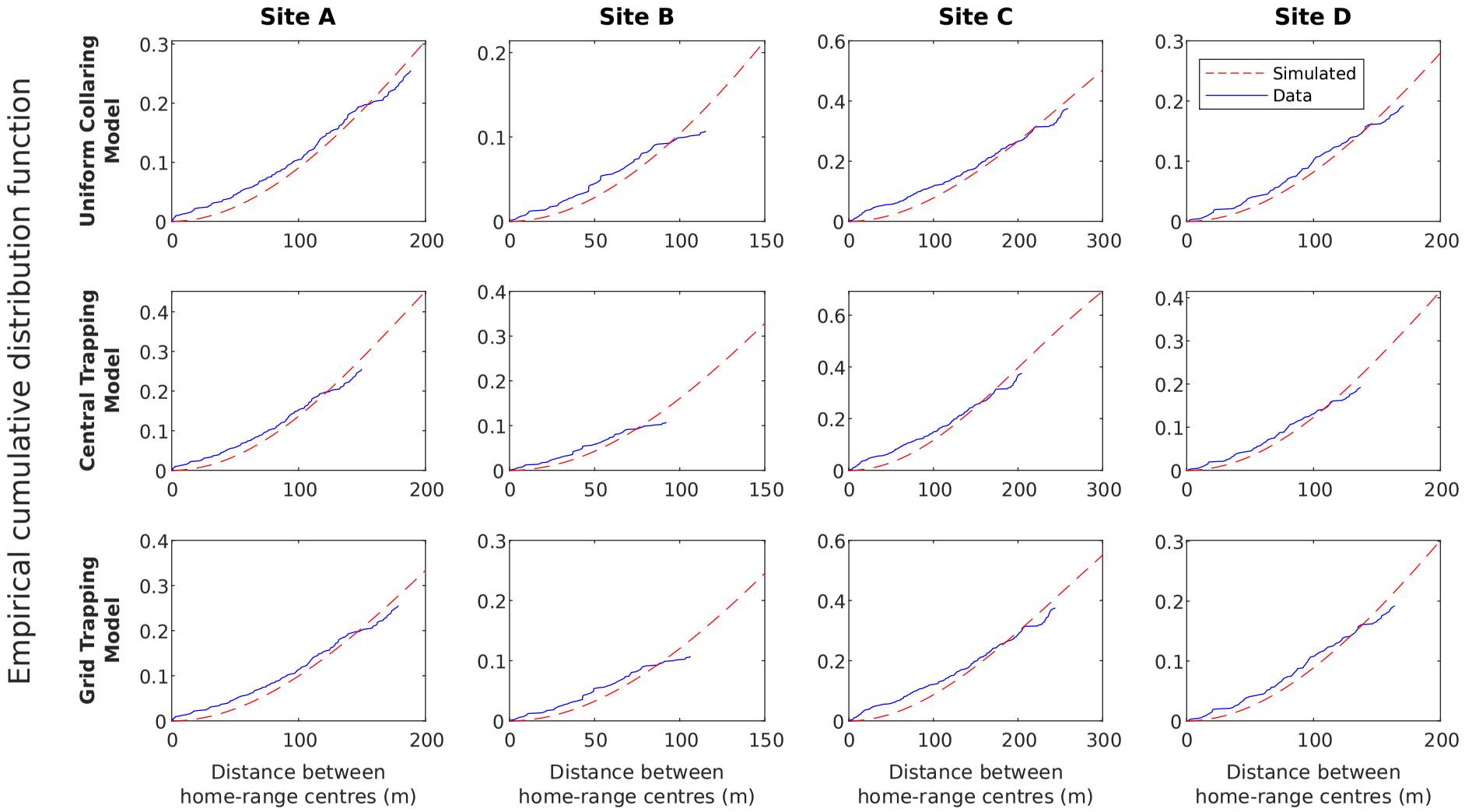



Figure 3.12: Comparison between the eCDFs of idealised model populations of $C \times 10^4$ individuals (red) and estimated distances between possum pairs in the data (blue), produced using the mean optimal value of the diffusion coefficient given in Table 3.2.

### 3.5.3 Home-range size prediction

We investigate our model's ability to predict the size of an individual's home-range. We define an individual's home-range to be the area that, on average, 95% of their movements during any given night lie within. Each individual's home-range is assumed to be identical and circular. We calculate the home-range sizes of possums in the data, and compare this with home-range sizes calculated using our estimated diffusion coefficients.

Using estimated diffusion coefficients, we calculate home-range areas for possums in each of our four possum distribution models by first assuming that each individual has a circular, bivariate normally distributed area of activity. This bivariate normal distribution has an associated covariance matrix $(\mathcal{D}t_{max})I_2$, where $t_{max} = 12$ hours. We calculate the radius $r$ such that $\mathbb{P}(Z \leq r^2) = 0.95$, for $Z \sim N_2(\mathbf{0}, (\mathcal{D}t_{max})I_2)$. Using the relationship between the bivariate normal and Rayleigh distributions, given in Fletcher (2017), we have that this radius is calculated as $r = \sqrt{-2(\mathcal{D}t_{max})\ln(1-0.95)}$. Hence, for a diffusion coefficient $\mathcal{D}$, the circular home-range area is given by $-2\pi(\mathcal{D}t_{max})\ln(1-0.95)$.

Richardson et al. (2017) provides estimated sizes of home-range areas, using trapping-location data. During a field study conducted between April and November 2012, 608 possums were trapped on four non-contiguous field sites. On each site, a square grid of 100 baited traps placed at 40m intervals was used. The locations, time, and possum details of each of the 7697 trapping events was recorded. A detailed description of how these trapping location data were collected is given in Section 2.1. Home-range size estimates were calculated for possums trapped on ten or more occasions, by drawing convex polygons that encompassed the locations of their trapping events. This approach resulted in a mean (across all the possums) home-range area of 1.52ha.

Our model does not describe the trapping of individuals during their nocturnal movements. Therefore, the approach taken in Richardson et al. (2017) cannot be directly compared to the results from our model. Instead, we use the approach from Efford (2004). Here a possum's activity range is assumed to be circular and bivariate normally distributed, with covariance matrix $\sigma^2 I_2$, where $\sigma$ is the spatial scale parameter. An estimate of the home-range radius is then calculated as $\sqrt{-2\sigma^2\ln(1-0.95)} \approx 2.45\sigma$. Squaring and multiplication by $\pi$ then gives the home-range area. Using values of $\sigma$ calculated in Richardson et al. (2017) a mean home-range area (across the entire possum corpus) of 4.46ha is calculated.

We compare home-range area estimations for each site from $\sigma$ (mean $\pm$ s.e., see Table 3.2), with estimations from each of our four possum distribution models, using $\mathcal{D}$ (mean $\pm$ s.d., see Table 3.2). Fig. 3.13 shows these estimations. For sites A and C, no model predicts the home-range size well. For site B, the Central Trapping Model gives a range with the closest mean and standard deviation to that of the data. The same holds for site D and the Grid Trapping Model. The overestimation of the home-range sizes for sites A, B and C is most likely due to the assumption that our random walkers follow an unbiased random walk. Due to the lack of tendency towards a particular point, especially their home-range centre, the individuals in our model are moving around in a much larger area than is realistic.

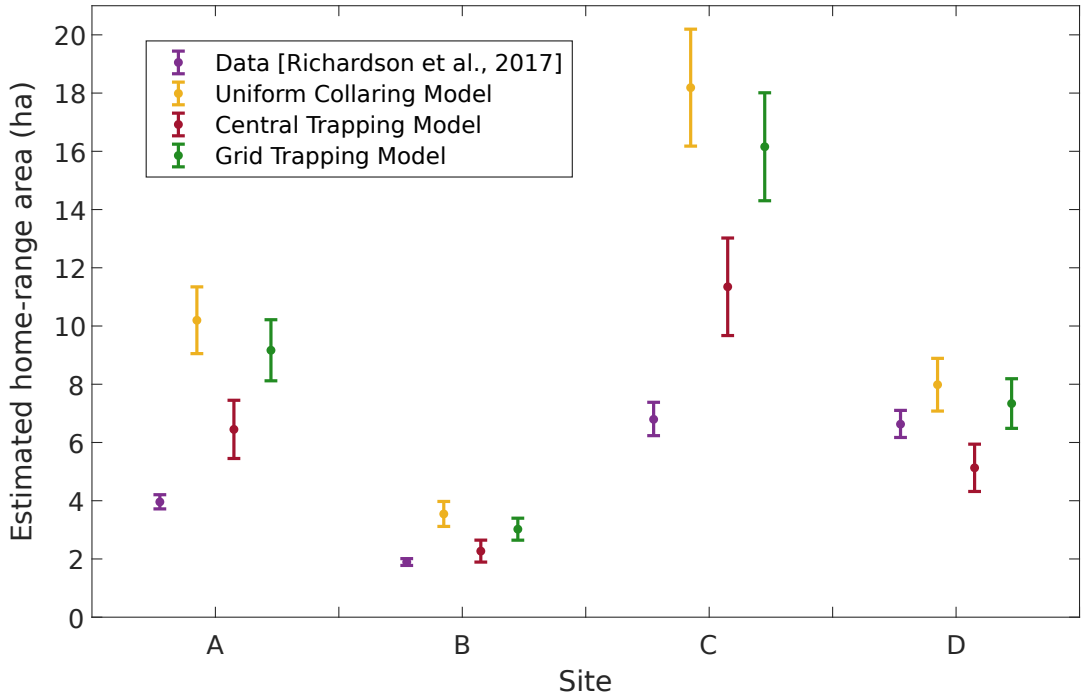



Figure 3.13: **Comparison of estimated ranges of home-range area, from data [Richardson et al., 2017], and from our possum distribution models.** Home-range areas calculated from values of $\sigma$ (mean $\pm$ s.e.) drawn from Richardson et al. (2017) are compared with ranges of home-range area calculated using the distributions of $\mathcal{D}$ (mean $\pm$ 1 s.d.) from each possum distribution model (see Table 3.2). For sites B and D, the Uniform Collaring Model gives the closest results. For sites A and C, no model performs well.

### 3.5.4 Summary

We have presented a method for estimating a diffusion coefficient for our individual-based model of possum movement. We produce a set of pairwise distances between individuals in a model population, and then compare the distribution of those distances with the distribution of estimated distances between possum pairs in our data set. The comparison is done by calculating the mean squared error (MSE) between the empirical cumulative distribution functions (eCDFs) of the model and data distances. We then find a value or range of values of $\mathcal{D}$ that minimise the MSE.

Distributions of estimated diffusion coefficients for each site and possum collaring model were then produced. The Uniform Collaring Model consistently gives a higher mean optimal diffusion coefficient and standard deviation compared with the Grid Trapping Model and the Central Trapping Model. We showed that these estimated diffusion coefficients give sensible percentages of nocturnal time spent by individuals traveling. Further, a comparison was made between the eCDFs of idealised populations and of pairwise distances between possum pairs in the data, calculated using the mean optimal diffusion coefficient. This comparison showed that, for the Central Trapping Model, the mean squared error calculated is consistently larger than when using either the Grid Trapping Model or the Uniform Collaring Model, the latter of which consistently gives the smallest MSE.

Finally, we estimated home-range sizes from both our estimated diffusion coefficients, and population-relevant parameters drawn from Richardson et al. (2017). We showed that for sites A and C, no possum collaring model replicates the home-range size well, whereas for sites B and D, the Central Trapping Model and the Grid Trapping Model respectively provide ranges closest to that of the data. The discrepancies and overestimation of the home-range sizes is most likely a result of the use of an unbiased random walk, resulting in individuals travelling in a larger environment than is realistic.

Having now estimated diffusion coefficients for our individual-based model, we go onto produce binary and weighted networks from the model, and then compare network characteristics between our model networks and empirical networks produced from the data.

## 3.6 Constructing binary contact networks from model simulations

For a population of $N$ individuals, a binary contact network $G$ is constructed by first assuming each individual, positioned at their home-range centre, is a vertex. The home-ranges centres are distributed as per Section 3.4. Using pairwise distances and a diffusion coefficient $\mathcal{D}$, values of the contact measure $P_{ij}$, discussed in Section 3.3.1, are calculated. Edges exist between individuals if $P_{ij}$ is greater than or equal to a given threshold value $\theta$.

We investigate the validity of our home-range centre distribution model and our possum collaring models by comparing certain network characteristics between binary networks produced from each model, and empirical binary contact networks produced from data (illustrated in Fig. 2.5). For each model, $10^4$ binary networks are produced. Home-range centres are distributed as per the model given in Section 3.4.1, and possums are collared according to the collaring models described in Section 3.4.2. The mean value, $\mu_{\mathcal{D}}$, from the estimated diffusion coefficient distributions (Table 3.2), is used to calculate the values of $P_{ij}$ for each pair. A threshold value of $\theta = 1/190$ is used. This value is used because it replicates the value of the probability $P_{ij}$ such that the expected number of nights that a contact occurs, during an observation period of 190 nights, is at least one.

### 3.6.1 Results

We compare five metrics of population network structure between the data and our models: network density, number of 3-cycles, transitivity coefficient, number of components, and the size of the giant component. A summary of this comparison is given in Table 3.5, and illustrated by boxplots in Fig. 3.14. The interquartile ranges are shown as the coloured boxes, with the whiskers going from the minimum to the maximum values.

The number of components and giant component size are best estimated by the Uniform Collaring Model on every site except D, where a better estimated is given by the Central Trapping Model. There is a negative correlation between mean diffusion coefficient and number of components. As the diffusion coefficient increases, the threshold distance will also increase, which will connect more individuals and result in fewer components.

Each of the network density, number of length-3-cycles, and the transitivity coefficient, are overestimated by our models in comparison with the empirical values. This could possibly be caused by our possum movement and contact model not accurately reflecting the tendency of possums to remain solitary except for during the mating season. As such, the number of possum pairs that come into contact, and thereby the network density, will be overestimated. The Uniform Collaring Model, across every site, gives the closest mean network density to that of the empirical contact networks (Table 3.5). As illustrated in Fig. 3.14A, the network density is consistently lowest in that model, with the Grid Trapping Model and the Central Trapping Model having larger values, always in that order. This order is most likely a result of distribution methods of the possums. The Uniform Collaring Model is the most spread out, resulting in a higher network density. Looking then to the Grid Trapping Model and the Central Trapping Model, each model's home-range centres are progressively more concentrated about various points.

Similarly, the propensity of pairwise connections to form between groups of three might also be overestimated, hence why the number of length-3-cycles is not well estimated on any site, and the number is consistently higher in our model populations than in the empirical data. Alternatively, it could instead be a result of the possums in the data being actually clustered, meaning that they are likely to come into contact with others in their cluster and form 3-cycles. Overall, the Uniform Collaring Model produces the closest mean value, across all sites, to the empirical value (Table 3.5 and Fig. 3.14B).

The transitivity coefficient is always overestimated (Table 3.5). As illustrated by Fig. 3.14C, the transitivity coefficients of the Central Trapping Model, the Grid Trapping Model, and the Uniform Collaring Model are similar, which is a result of the similarity in the distribution of their home-range centres. However, the closest mean transitivity coefficient to the empirical value is produced using Central Trapping Model in the cases of sites A, B, and D. For site C, the Grid Trapping Model produced the closest mean value to the data.

| Site | Model | Network size | Components | Giant component size | Network density | Length-3-cycles | Transitivity coefficient |
|---|---|---|---|---|---|---|---|
| A | Data | 39 | 1 | 39 | 0.2551 | 359 | 0.1841 |
| | Uniform Collaring | | 1.0035 | 38.99448 | 0.29986 | 582.1911 | 0.2209 |
| | Central Trapping | | 1.0915 | 38.88866 | 0.30652 | 652.8414 | 0.2197 |
| | Grid Trapping | | 1.0053 | 38.99150 | 0.30073 | 588.2659 | 0.2209 |
| B | Data | 41 | 1 | 41 | 0.1073 | 50 | 0.1362 |
| | Uniform Collaring | | 1.7065 | 38.41539 | 0.13766 | 133.1555 | 0.2080 |
| | Central Trapping | | 3.0835 | 37.29201 | 0.14050 | 154.5003 | 0.2043 |
| | Grid Trapping | | 1.7950 | 38.30326 | 0.13770 | 134.0186 | 0.2076 |
| C | Data | 40 | 1 | 40 | 0.3756 | 943 | 0.2028 |
| | Uniform Collaring | | 1.0000 | 39.99999 | 0.41663 | 1274.8661 | 0.2310 |
| | Central Trapping | | 1.0051 | 39.99436 | 0.42307 | 1387.6783 | 0.2321 |
| | Grid Trapping | | 1.0001 | 39.99991 | 0.41694 | 1281.4251 | 0.2308 |
| D | Data | 39 | 3 | 37 | 0.1922 | 167 | 0.1745 |
| | Uniform Collaring | | 1.0469 | 38.89118 | 0.22351 | 313.0605 | 0.2148 |
| | Central Trapping | | 1.4052 | 38.44159 | 0.22843 | 354.8133 | 0.2121 |
| | Grid Trapping | | 1.0594 | 38.87399 | 0.22416 | 316.0954 | 0.2144 |

Table 3.5: **Comparison of summary statistics from empirical contact networks and from binary networks produced from each of the possum distribution models.**

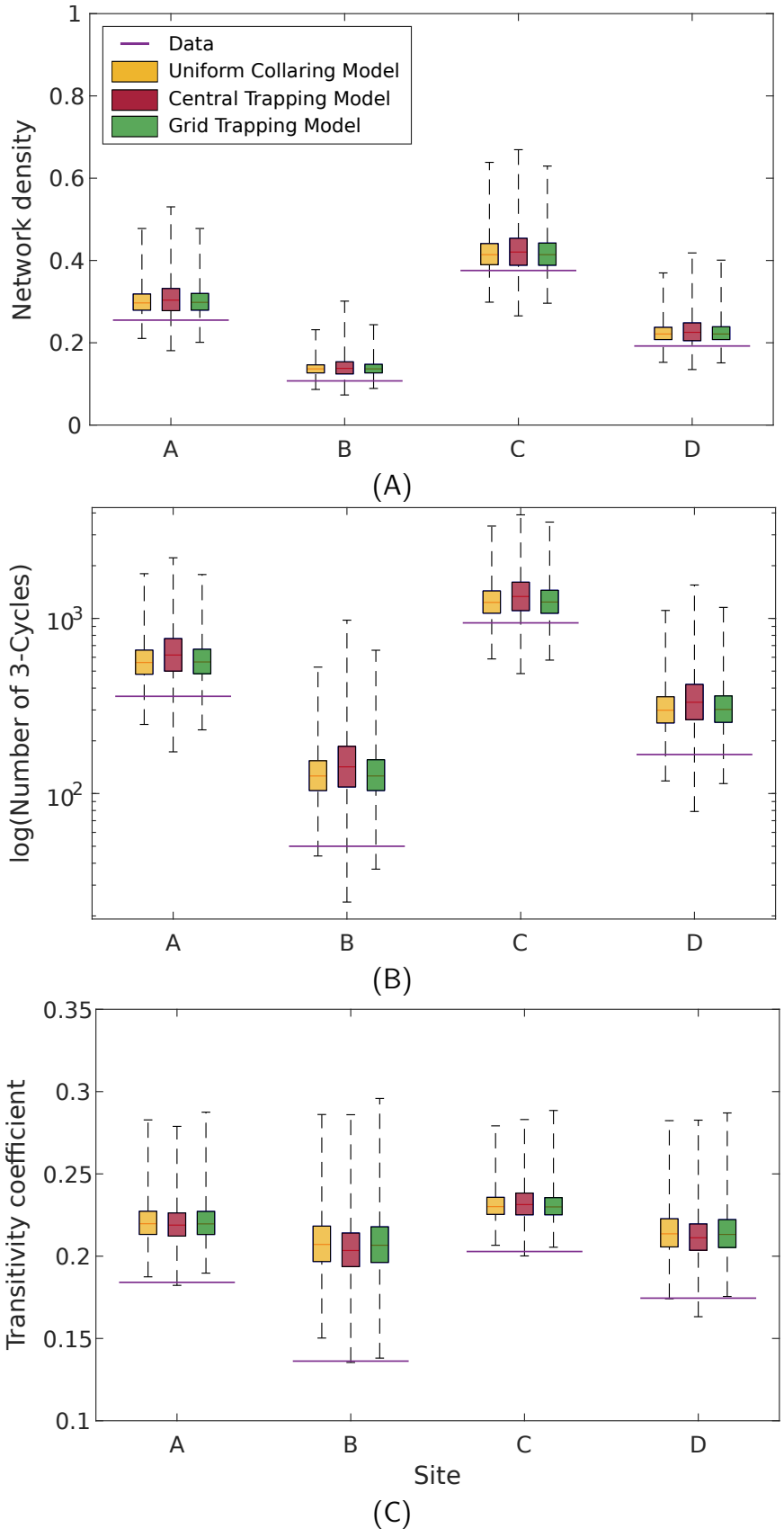
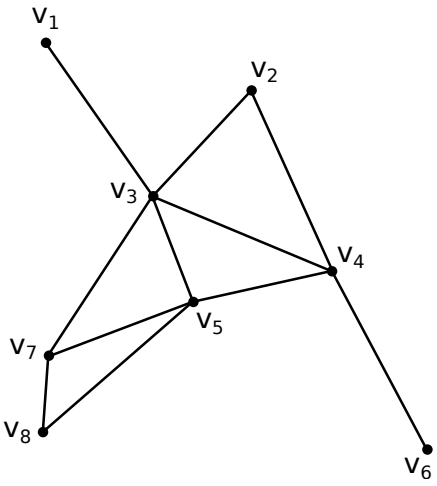



Figure 3.14: **Box plots for the network density, number of 3-cycles, and transitivity coefficient produced from the various models.** The interquartile ranges are shown as the coloured boxes, with the whiskers going from the minimum to the maximum values. The network density is well-estimated (Fig. 3.14A), as is the number of 3-cycles, with the exception of site B (Fig. 3.14B), but the transitivity coefficient is not well estimated on any site (Fig. 3.14C).

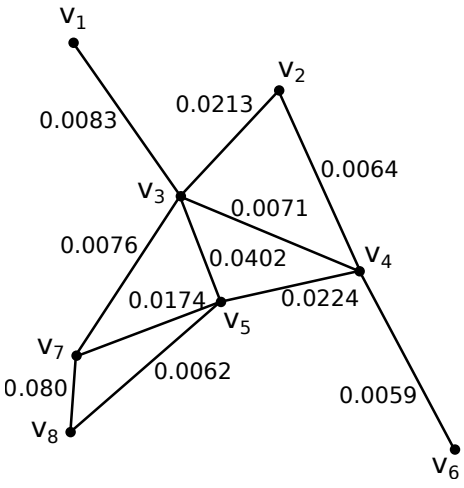
## 3.7 Constructing weighted contact networks from model simulations

From each possum collaring model, three weighted contact networks are produced. The underlying binary network of all three is produced as per Section 3.6. The edge weightings are calculated using the pairwise distances $\hat{d}_{ij}$ between nodes, the mean diffusion coefficient for each model given in Table 3.2, and the measures of contact length and importance shown in Fig. 3.2.

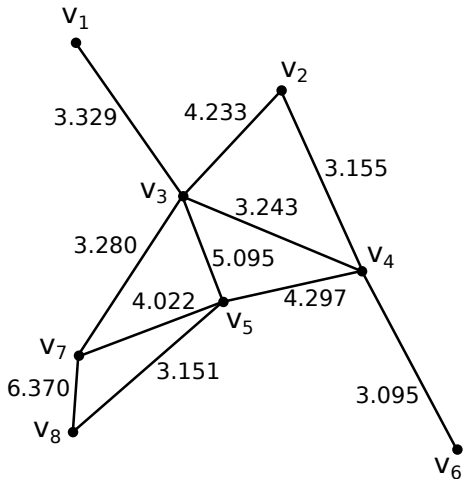
Two of the weighted networks are weighted with, respectively, the contact probability $P_{ij}$ and the expected total time spent in contact $T_{ij}$. A higher value of $P_{ij}$ or $T_{ij}$ indicates a higher importance of the contacts between pairs of individuals. However, the opposite is true for $F_{ij}$, with a higher value of the first contact time indicating a lower contact importance. Therefore, networks weighted with $P_{ij}$ or $T_{ij}$ cannot directly be compared with networks weighted with $F_{ij}$. To address this, the edges of the third weighted contact network are weighted with $F_{\max} - F_{ij}$, as a higher value of $F_{\max} - F_{ij}$ indicates a higher contact importance. Here $F_{\max} = 12$ hours is the maximal first contact time. A toy example of this construction process is shown in Fig. 3.15.



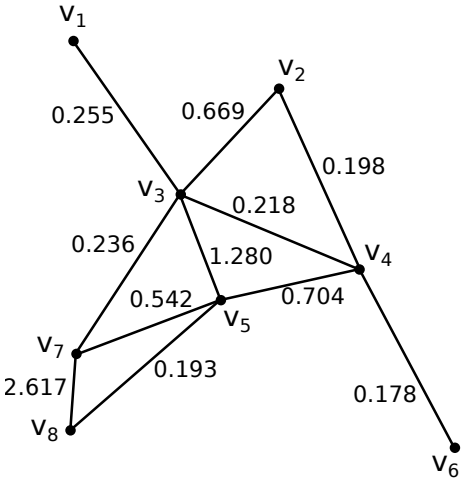

(A) Binary contact network $G$




(B) Network weighted by $P_{ij}$.




(C) Network weighted by $F_{\max} - F_{ij}$ (hrs).




(D) Network weighted by $T_{ij}$ (s).

Figure 3.15: **Toy example of the construction of weighted contact networks from model simulations.** Vertices, representing individuals, are embedded in $\mathbb{R}^2$ at their home-range centres. Contact importance measures, described in Section 3.3.1, are calculated using pairwise distances and $\mathcal{D} = 0.1\text{m}^2\text{s}^{-1}$. Edge exist between vertices if the contact probability $P_{ij} \geq 1/190$. We can produce three weighted contact networks from our initial binary network, weighted using the three contact importance measures.

### 3.7.1 Results

To assess the validity of each of our possum collaring models, we compare the distributions of edge weights between weighted contact networks produced from these models and empirical weighted networks from the data. We produce multiple weighted contact networks and, for each network, use the two-sample Kolmogorov-Smirnov test to test the null hypothesis, $H_0$, that the edge weights of our simulated contact network are from the same distribution as the empirical edge weights.

Summarised in Tables 3.6 to 3.8 are the results of performing the Kolmogorov-Smirnov test. In these table are the mean Kolmogorov–Smirnov test statistic (with its standard deviation in brackets), the mean asymptotic $p$-value, and the proportion of realisations where the null hypothesis $H_0$ was rejected. From these large mean test statistics, low $p$-values, and consistent rejection of $H_0$, we can conclude that our weighted network models do not replicate well the distributions of $F_{\max} - F_{ij}$ and $T_{ij}$ edge weights. When considering these results in the case of the $P_{ij}$ edge weights, our models perform comparatively better.

| $P_{ij}$ | | | | |
|---|---|---|---|---|
| Site | Model | Mean test statistic | Mean asymptotic $p$-value | Proportion of realisations where $H_0$ is rejected |
| A | Uniform Collaring | 0.1701 (0.0322) | 0.0268 | 0.8668 |
| | Central Trapping | 0.1832 (0.0310) | 0.0113 | 0.9494 |
| | Grid Trapping | 0.1707 (0.0321) | 0.0257 | 0.8727 |
| B | Uniform Collaring | 0.2979 (0.0426) | 0.0030 | 0.9900 |
| | Central Trapping | 0.3111 (0.0413) | 0.0014 | 0.9965 |
| | Grid Trapping | 0.2996 (0.0422) | 0.0026 | 0.9917 |
| C | Uniform Collaring | 0.1432 (0.0186) | 0.0103 | 0.9550 |
| | Central Trapping | 0.1536 (0.0182) | 0.0040 | 0.9916 |
| | Grid Trapping | 0.1428 (0.0180) | 0.0090 | 0.9650 |
| D | Uniform Collaring | 0.1837 (0.0366) | 0.0502 | 0.7613 |
| | Central Trapping | 0.1985 (0.0352) | 0.0250 | 0.8786 |
| | Grid Trapping | 0.1853 (0.0367) | 0.0470 | 0.7771 |

Table 3.6: **The $P_{ij}$ edge weight distributions are best replicated by the Uniform Collaring Model.** Compared with the distributions of $F_{\max} - F_{ij}$ and $T_{ij}$, the $P_{ij}$ edge weight distributions are replicated better.

| $F_{max} - F_{ij}$ | | | | |
|---|---|---|---|---|
| Site | Model | Mean test statistic | Mean asymptotic $p$-value | Proportion of realisations where $H_0$ is rejected |
| A | Uniform Collaring | 0.7057 (0.0244) | $2.52\times10^{-31}$ | 1 |
| | Central Trapping | 0.7168 (0.0228) | $6.26\times10^{-34}$ | 1 |
| | Grid Trapping | 0.7069 (0.0241) | $6.25\times10^{-33}$ | 1 |
| B | Uniform Collaring | 0.8064 (0.0264) | $2.20\times10^{-21}$ | 1 |
| | Central Trapping | 0.8135 (0.0254) | $5.18\times10^{-22}$ | 1 |
| | Grid Trapping | 0.8079 (0.0261) | $1.33\times10^{-22}$ | 1 |
| C | Uniform Collaring | 0.7859 (0.0190) | $2.03\times10^{-58}$ | 1 |
| | Central Trapping | 0.7957 (0.0172) | $3.80\times10^{-65}$ | 1 |
| | Grid Trapping | 0.7866 (0.0184) | $5.07\times10^{-59}$ | 1 |
| D | Uniform Collaring | 0.7838 (0.0253) | $1.85\times10^{-28}$ | 1 |
| | Central Trapping | 0.7933 (0.0237) | $3.74\times10^{-30}$ | 1 |
| | Grid Trapping | 0.7856 (0.0250) | $2.99\times10^{-29}$ | 1 |

Table 3.7: **No model replicates the $F_{max} - F_{ij}$ edge weight distribution well, but the Uniform Collaring Model produces the lowest mean Kolmogorov–Smirnov test statistic, and the largest asymptotic $p$-values.**

| $T_{ij}$ | | | | |
|---|---|---|---|---|
| Site | Model | Mean test statistic | Mean asymptotic $p$-value | Proportion of realisations where $H_0$ is rejected |
| A | Uniform Collaring | 0.8129 (0.0163) | $3.27\times10^{-48}$ | 1 |
| | Central Trapping | 0.7533 (0.0182) | $5.97\times10^{-41}$ | 1 |
| | Grid Trapping | 0.7992 (0.0167) | $5.18\times10^{-46}$ | 1 |
| B | Uniform Collaring | 0.6179 (0.0338) | $3.69\times10^{-13}$ | 1 |
| | Central Trapping | 0.5330 (0.0311) | $5.03\times10^{-10}$ | 1 |
| | Grid Trapping | 0.5851 (0.0327) | $3.50\times10^{-12}$ | 1 |
| C | Uniform Collaring | 0.8971 (0.0101) | $2.58\times10^{-87}$ | 1 |
| | Central Trapping | 0.8472 (0.0151) | $8.05\times10^{-76}$ | 1 |
| | Grid Trapping | 0.8866 (0.0112) | $3.38\times10^{-86}$ | 1 |
| D | Uniform Collaring | 0.6821 (0.0279) | $1.06\times10^{-20}$ | 1 |
| | Central Trapping | 0.5983 (0.0314) | $9.63\times10^{-16}$ | 1 |
| | Grid Trapping | 0.6670 (0.0288) | $1.26\times10^{-19}$ | 1 |

Table 3.8: **No model replicates the $T_{ij}$ edge weight distribution well, but the Central Trapping Model produces the lowest mean Kolmogorov–Smirnov test statistic, and the largest asymptotic $p$-values.**

We consider the distributions of the Kolmogorov–Smirnov test statistic, defined in Section 3.2.3, which are illustrated for site D in Fig. 3.16. We can conclude that, when considering the networks weighted with $P_{ij}$ and $F_{\max} - F_{ij}$, the Uniform Collaring Model and the Grid Trapping Model produce edge weight distributions with a lower test statistic than those produced using the Central Trapping Model. The opposite is true for networks weighted with $T_{ij}$, where the Central Trapping Model outperforms the Grid Trapping Model, which in turn outperforms the Uniform Collaring Model.

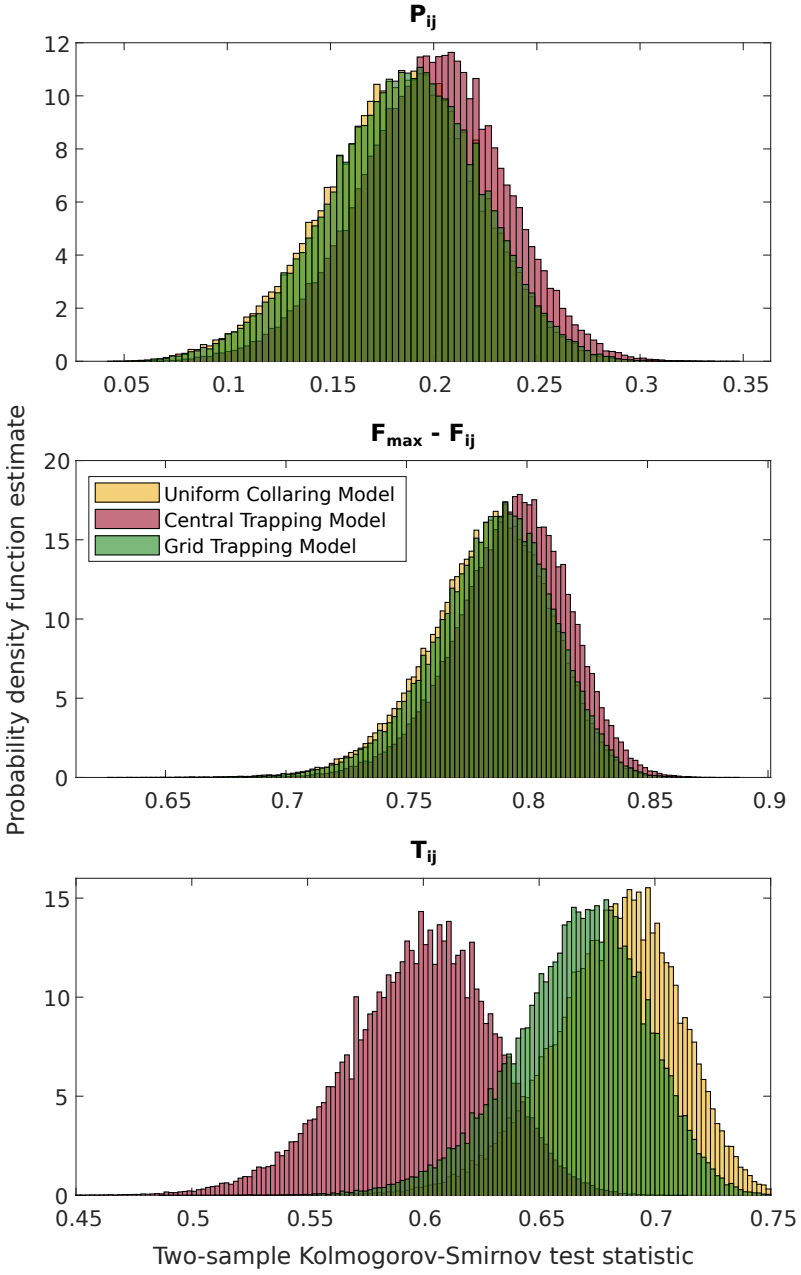



Figure 3.16: **Distributions of Kolmogorov–Smirnov test statistics produced by performing the two-sample Kolmogorov–Smirnov test on edge weight distributions produced from model networks, and from the empirical data networks.** For networks weighted with $P_{ij}$ and $F_{\max} - F_{ij}$, the Uniform Collaring Model and the Grid Trapping Model produce edge weight distributions with a lower test statistic than those produced using the Central Trapping Model. The opposite is true for networks weighted with $T_{ij}$.

## 3.8 Discussion

In this chapter we presented a model that aimed to describe the nocturnal movements of and contacts between members of a possum population.

From our movement model, we can produce simulated and analytical values of three contact measures. The deterministic relationships, shown in the simulated values, between these measures are not also shown in empirical values from the data, with the exception of the contact probability $P_{ij}$ and total contact time $T_{ij}$. This could indicate that our movement model is not replicating some aspect of possum behaviour. Two possible aspects come to mind. The first is the 'memory-less' property that all three measures exhibit. Our model assumes that events on previous nights do not affect future nights. This is unrealistic as possums may have preferred locations for resources, and may avoid other locations because of terrain or other possums. The second is the assumption that our model make, that behaviour subsequent to a contact is not affected by the contact, which again is unrealistic. For example, if two possums meet and then fight, their subsequent movements are likely to be in different directions, as to avoid meeting again.

Due to a lack of spatial information in the data set we compare our results to, we have described three different models for the distribution of $C$ collared possum home-range centres. We use each of these models to estimate values of the diffusion coefficient $\mathcal{D}$ and home-range areas. On each site, the Uniform Collaring Model provides diffusion coefficients that give the best fit between an idealised population and the data. When examining the estimated home-range area, the Uniform Collaring Model performs best for sites B and D, whereas no model performs well for sites A and site C.

We produce binary networks representing our populations, using mean estimated diffusion coefficients $\mu_{\mathcal{D}}$ and the contact measure $P_{ij}$. When comparing with empirical contact networks from data, we examine the network density, number of 3-cycles, transitivity coefficient, number of components, and the size of the giant component. No one model performs well when estimating each network metric, but the Uniform Collaring Model performs best overall.

Summarised in Tables 3.6 to 3.8 are the results of performing the two-sample Kolmogorov-Smirnov test. In these table are the mean Kolmogorov–Smirnov test statistic (with its standard deviation in brackets),

the mean asymptotic $p$-value, and the proportion of realisations where the null hypothesis $H_0$ was rejected. From these large mean test statistics, low $p$-values, and consistent rejection of $H_0$, we can conclude that our weighted network models do not replicate well the distributions of $F_{\max} - F_{ij}$ and $T_{ij}$ edge weights. When considering these results in the case of the $P_{ij}$ edge weights, our models perform comparatively better.

Finally, weighted contact networks are also produced. We perform the two-sample Kolmogorov-Smirnov test to test the null hypothesis, $H_0$, that the edge weights of our simulated contact network are from the same distribution as the empirical edge weights. Our weighted network models do not replicate the $F_{\max} - F_{ij}$ and $T_{ij}$ edge weight distributions well, but do replicate the $P_{ij}$ distribution comparatively better. Overall, for $P_{ij}$ and $F_{\max} - F_{ij}$, the Uniform Collaring Model produces the lowest mean Kolmogorov-Smirnov test statistic and highest asymptotic $p$-values. The same holds for $T_{ij}$ and the Central Trapping Model.

In conclusion, the Uniform Collaring Model performs best for all sites. A summary of this is given in Table 3.9.

| | Site | | | |
|---|---|---|---|---|
| Metric | A | B | C | D |
| Mean squared error for $\mathcal{D}$ (Table 3.2) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Uniform Collaring |
| Home-range area estimation (Fig. 3.13) | - | Uniform Collaring | - | Uniform Collaring |
| Network density (Fig. 3.14A) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Uniform Collaring |
| Length-3-cycles (Fig. 3.14B) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Uniform Collaring |
| Transitivity coefficient (Fig. 3.14C) | Central Trapping | Central Trapping | Grid Trapping | Central Trapping |
| Components (Table 3.5) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Central Trapping |
| Giant component size (Table 3.5) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Central Trapping |
| Kolmogorov–Smirnov test for $P_{ij}$ (Table 3.6) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Uniform Collaring |
| Kolmogorov–Smirnov test for $F_{\max} - F_{ij}$ (Table 3.7) | Uniform Collaring | Uniform Collaring | Uniform Collaring | Uniform Collaring |
| Kolmogorov–Smirnov test for $T_{ij}$ (Table 3.8) | Central Trapping | Central Trapping | Central Trapping | Central Trapping |

Table 3.9: **The Uniform Collaring Model performs best for all sites.**

## 3.9 Glossary of notation used

| Notation | Meaning |
|---|---|
| $i$, $j$ | Individuals in our population. |
| $\mathbf{z}_i$ | Centre of an individual $i$'s home-range. |
| $r$ | Home-range radius (m). |
| $\mathbf{z}_{ij}$ | Vector difference between the home-range centres of $i$ and $j$. $\mathbf{z}_{ij} = \mathbf{z}_{ij} - \mathbf{z}_{ij}$. |
| $d_{ij}$ | Distance between the home-ranges centres of $i$ and $j$. $d_{ij} = \|\mathbf{z}_{ij}\|_2$. |
| $p$ | Probability of moving on the lattice from the current position, in any direction. |
| $\delta$ | Spacing of the regular square lattice. |
| $\tau$ | Time step. |
| $\mathcal{D}$ | Diffusion coefficient ($\mathrm{m^2 s^{-1}}$). $\mathcal{D} = p\delta^2/2\tau$. |
| $\mu_{\mathcal{D}}$ | Mean of the diffusion coefficient ($\mathrm{m^2 s^{-1}}$). |
| $\mathbf{x}_t$ | Position vector of an individual at time $t$. $\mathbf{x}_0 = \mathbf{z}_i$. |
| $\mathbf{w}_t$ | Vector difference between a pair of individuals at time $t$. $\mathbf{w}_0 = \mathbf{z}_{ij}$. |
| $\Delta_t$ | Scaled Euclidean distance between $i$ and $j$ at time $t$. $D_t = \|\mathbf{w}_t\|_2/\sqrt{2\mathcal{D}t}$. |
| $P_{ij}$ | Probability that at least one contact will occur between $i$ and $j$ during a given night. |
| $F_{ij}$ | Expected nightly first contact time between $i$ and $j$, given a contact occurs. |
| $T_{ij}$ | Expected total time $i$ and $j$ spent in contact during a given night. |
| $t_{\max}$ | Variable equivalent to 12 hours. |
| $F_{\max}$ | Maximal first contact time. |
| A, B, C, D | Study site designations. |
| $N$ | Number of (trapped) possums. |
| $C$ | Number of possums collared on a given study site. |
| $\omega$ | Proportion of possible possum pairs that came into contact. |
| $\lambda$ | Density parameter of the spatial Poisson point process. $\lambda > 0$. |
| $R_I$ | Radius of the area model individuals' home-range centres lie in (m). $R_I = \sqrt{N/\pi\lambda}$. |
| $c_i$ | Probability of an individual $i$ being trapped during the observation period. |
| $T_{\text{central}}$, $T_{\text{grid}}$ | Functions describing the probability $q_i$ of being trapped during the observation period. |
| $n_{\mathbf{traps}}$ | Number of traps. $n_{\mathbf{traps}} = 100$. |
| $\mathbf{y}_j$ | Location of the $j^{th}$ trap. |
| $g_0$ | One-night probability of capture at the home-range centre. |
| $\sigma$ | Measure of home-range size (m). $r = \sqrt{-2\sigma^2 \ln 1 - 0.95}$. |
| $d_{model}$ | Pairwise distances between home-range centres of model individuals. |
| $d_{data}$ | Estimated pairwise distances between home-range centres of possum pairs in the data. |
| $U_{model}$, $U_{data}$ | Unique values of $d_{model}$ and $d_{data}$ respectively. |
| $f_{model}$, $f_{data}$ | Empirical cumulative distribution functions of $d_{model}$ and $d_{data}$ respectively. |

Table 3.10: Table of notation

# Chapter 4

# A stochastic transmission model for the spread of an epidemic through a network

## 4.1 Background

In Chapter 2, we discussed data regarding the experimental infections of possums in a population, and the subsequent secondary infections. In order to model this effectively, which we do in Chapter 5, we present in this chapter a stochastic transmission model for the simulation of an epidemic through a population. Our model utilises a weighted contact network representation of the population to inform the disease spread.

By performing simulations on toy networks, we then investigate the effect on epidemic spread of changing aspects of the epidemic, including the disease incubation period and the likelihood of a contact resulting in infection transmission, in addition to network topology. We find that the biggest effect can be seen when altering either the parameters determining how the epidemic progresses, or the network density. This transmission model will be used in the next chapter to simulate bovine tuberculosis epidemics spreading on the weighted contact networks produced from the contact model presented in Chapter 3.

The dynamics of disease spread will be modelled using a compartmental epidemic model, specifically an SEIR-model. Here, the population is compartmentalised according to what stage of the epidemic individuals

are currently at. Individuals can move between compartments according to certain parameters. For example, in the SEIR-model, the population is divided into susceptible (meaning not currently or previously infected) individuals, who have a probability per unit time of being exposed to the disease by any of their currently infectious neighbours in the contact network. This probability increases with the number of infectious individuals. Exposed individuals are initially noninfectious, but become so after a latency period. These now infectious individuals are removed from the population after a period of time.

Compartmental models date back to the 1927 paper by Kermack and McKendrick. Many different processes have been modelled using a compartmental framework, including predator-prey models (see [Kohlmeier and Ebenhöh, 1995, Pitchford and Brindley, 1998, Wang et al., 2011] and references within), interactions between combatants during warfare (see [Ekchian, 1982, Johnson and MacKay, 2011, Armstrong, 2014] and references within), chemical and biological reactions (see [Zhao et al., 2017, Darvey and Staff, 1966, Erban and Chapman, 2009] and references within), and even revolutions in dictatorial regimes [Lang and De Sterck, 2014].

Compartmental models have also been produced to describe the spread of various epidemics through a variety of different species, including animal, plant and insect. Humans diseases in particular have been modelled extensively. For example, compartmental models have been developed to describe outbreaks of Ebola hemorrhagic fever. Lekone and Finkenstädt (2006) produced a model for the 1995 outbreak in the Democratic Republic of the Congo, and Getz et al. 2019 and others produced models for the 2014–2015 outbreak. Other disease modelled include Spanish flu (Mills et al. (2004)), whooping cough in Nebraska, United States (Ameri and Cooper (2019)), and cholera in East African refugee camps (Hailegiorgis and Crooks (2012)). Diseases in animals and plants are also of interest, and pose different problems, as well as extensions, to those that arise when modelling human populations. These obstacles include estimation of epidemic parameters such as the rate of infection, and the determination the transmission route(s). Examples of animal diseases modelled include the 2001 Foot and Mouth disease outbreak in the United Kingdom, which was modelled by both Keeling et al. (2001) and Ferguson et al. (2001), and avian influenza in Pennsylvania, United States (see Rorres et al. (2011)).

Contacts between individuals are not necessary uniform in frequency or length, and some individuals may never meet. Therefore, the dynamics of epidemic spread, particularly for small populations, are better modelled using a network framework, as we use in this chapter. The method of modelling an epidemic using a

social network is a common one [Kenah and Robins, 2007, Keeling and Eames, 2005, Newman, 2002]. This approach has been applied to many different diseases and contexts. Animal and plant diseases have been modelled, including devil facial tumour disease in Tasmanian devils [Hamede et al., 2009]. In their 2017 paper, Buddenhagen et al. built networks of the transfer of potato seeds between farmers, and simulated epidemics on those networks.

Human diseases have also been modelled by using a network. The spread of A/H1N1 (also known as Swine flu) has been modelled from a global perspective [Rausanu and Grosan, 2010], and in a single country, namely Mexico [Frias-Martinez et al., 2011]. Other diseases modelled include obesity [González-Parra et al., 2010], and the spread of dengue fever in Malaysia [Malik et al., 2017]. Weeden and Cornwell (2020) built networks of students on the campus of Cornell University, New York, United States, with an aim of providing insight into the spread of an epidemic such as COVID-19 through a unviersity population.

## 4.2 Setting

In this section we describe mathematical techniques that are necessary for the models of epidemics through a population that we showcase in this chapter.

### 4.2.1 Compartmental epidemic models

A *compartmental model* is a modelling technique where a population, be it of people, animals, chemical reactants, or gases, are divided into compartments or states. Members of a given compartment are assumed to be homogeneous with respect to some characteristic. In the case of epidemics, all individuals in the same compartment are assumed to be at the same stage or state of the epidemic being modelled. Members of the population can only inhabit one state at any given time, and have the ability to move between states at given rates.

We model an epidemic, spreading through a population of individuals, using a compartmental model. Individuals can occupy one of four states: Susceptible, Exposed (infected but not infectious), Infectious, and Removed (dead or recovered). The model representation of this epidemic is referred to as an SEIR-model. At each time $t$, the numbers of susceptible, exposed, infected, and removed individuals are given by $S(t)$, $E(t)$, $I(t)$, and $R(t)$ respectively. By assuming that the population is of size $N$, and that the rates of birth and non-disease-induced mortality are zero, the population size remains constant, and thus for each $t$, $S(t) + E(t) + I(t) + R(t) = N$.

The following assumptions are made:

1. Disease is transmitted, at a rate $\beta$, from an infectious individual to a susceptible individual by contact.
2. Exposed individuals become infectious at a constant rate $\gamma$.
3. Infectious individuals die, or are removed, from the population at a constant rate $\delta$. When an individual is removed, no further contacts with it can occur, thus it cannot transmit the infection to any other individuals.

These assumptions are illustrated in the flowchart shown in Fig. 4.1.

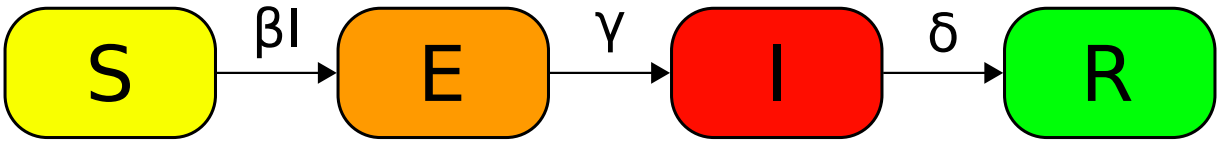



Figure 4.1: **Flowchart representing the state transitions in an SEIR-compartmental model.**

### 4.2.2 Basic reproductive number $R_0$

An important metric in epidemic modelling is the basic reproductive number $R_0$. First used in demography, where it was referred to as the 'net reproductive rate' and indicated the growth of a population [Dietz, 1993, Böckh, 1886], it started to come into use in an epidemiology context during the $20^{th}$ century, and is now widely used when studying the spread of disease [Heffernan et al., 2005]. A dimensionless variable, it is defined as the average or expected number of secondary infections produced by one infected individual in a completely susceptible population [Dietz, 1993, Fine, 1993, Diekmann et al., 1990, Heffernan et al., 2005].

Of particular importance to the study of epidemic spread are two features of $R_0$. The first is called the *epidemic threshold*, and refers to the threshold of $R_0$ about 1, where below 1 the disease will not be able to establish itself in a population, whereas above 1 it will. The second is that the proportion of a population that need to be vaccinated in order to stop an epidemic establishing itself is calculated as $1 - \frac{1}{R_0}$.

### 4.2.3 Gillespie algorithm

The Gillespie algorithm is a computational method used to simulate the evolution of a continuous time Markov process [Gillespie, 1977]. This method has been applied to a variety of different situations, including the movement of cells [Binny et al., 2016] and epidemics such as HIV/AIDS [Kamina et al., 2019].

At each time $t$, each individual $i$ in a population of size $N$ has at most one transition it can undergo, with an associated transition rate $q_i$. Therefore the system is updated by undergoing one of $N$ transitions. Some of these transition rates are 0. The process can be described in four steps:

1. Initialise the system by setting $t = 0$. The initial state of the system is $\mathbf{x}(0)$.

2. For the current system state $\mathbf{x}(t)$, the $N$ transition rates, $q_i(\mathbf{x}(t))$, are calculated as shown in Table 4.1. For ease of notation, we will denote $q_i(\mathbf{x}(t))$ by $q_i(t)$. The total transition rate $q(t)$ is then calculated as their sum, $q(t) = \sum_{i=1}^{N} q_i(t)$.

3. Two random numbers are chosen. The first, $\Delta t$, drawn from an exponential distribution with mean

$1/q(t)$, describes the time until the next event. The second is used to determine which of the transitions will occur, and is drawn from a uniform distribution on $(0, q(t)]$. If the second number lies in the interval $(q_{i-1}(t), q_i(t)]$, then the $i^{th}$ transition will occur, and the state of individual $i$ is changed accordingly. Here we define $q_0(t) = 0$.

4. Time is then progressed to $t + \Delta t$, the occurrence of the next transition, and the system state $x(t + \Delta t)$ is updated to reflect the transition that just took place.

Steps 2-4 are repeated until either a preset time limit is reached, or $q(t) = 0$.

#### 4.2.4 Random geometric graphs

A random geometric graph (RGG) is a network where the vertices are embedded in $\mathbb{R}^2$ and distributed according to a random process. An edge exists between nodes $i$ and $j$ if the euclidean distance between them is less than a certain threshold value. Random geometric graphs have been used to model networks of wireless sensors [Kenniche and Ravelomananana, 2010], protein–protein interactions [Higham et al., 2008], and virus spreading [Preciado and Jadbabaie, 2009].

## 4.3 Well-mixed Transmission Model

An important step in building our stochastic network transmission model is to examine the well-mixed epidemic model. We start off with the same population of size $N$ as in Section 4.2.1. In the well-mixed model, in addition to the three assumptions detailed in Section 4.2.1, an additional assumption is also made. This assumption is that the population is well-mixed, meaning an individual in the population makes contact with all other individuals at at given time. Therefore, the rate at which susceptible individuals become infected is proportional to the product of the numbers of susceptible and infectious individuals.

Approaching this from a deterministic viewpoint, the SEIR-model can be expressed in the form of a system of four ordinary differential equations, given in Eq. (4.1). A plot of the numerical solutions to this ODE system is shown in Fig. 4.2, with $N = 160$, $E(0) = 1$, $I(0) = R(0) = 0$, $\beta = 0.005$, $\gamma = 0.0133$, and $\delta = 0.0074$.

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dE}{dt} = \beta SI - \gamma E, \quad \frac{dI}{dt} = \gamma E - \delta I, \quad \frac{dR}{dt} = \delta I \tag{4.1}$$

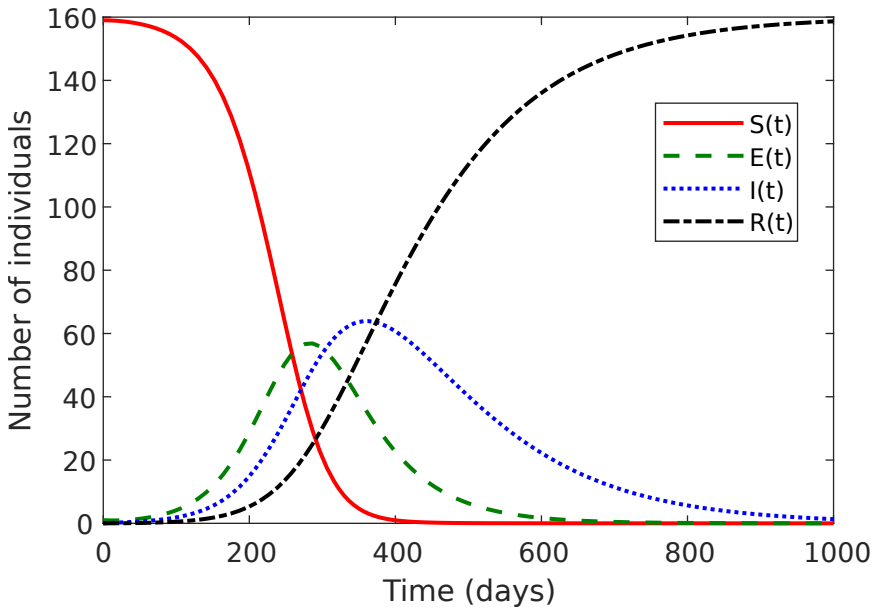



Figure 4.2: **Plot of the numerical solutions to the ordinary differential equation system given in Eq. (4.1), with $N = 160$, $E(0) = 1$, $I(0) = R(0) = 0$, $\beta = 0.005$, $\gamma = 0.0133$, and $\delta = 0.0074$.**

Let us examine assumptions (3) and (4) from Section 4.2.1, that individuals in states $E$ and $I$ leave that state at constant rates $\gamma$ and $\delta$ respectively. As explained in Bauer (2008), let $L(t)$ denote the number of exposed individuals who remained exposed but not infectious (i.e. latent) at time $t$ after they were infected.

Assumption (3), that exposed individuals become infectious at a constant rate $\gamma$, results in the relation $\frac{dL}{dt} = -\gamma L$ [Brauer, 2008]. The solution to this ODE is given by $L(t) = L(0)\exp(-\gamma t)$. Division by $L(0)$ shows that, at time $t$, the proportion of those exposed individuals who remain latent is $\exp(-\gamma t)$.

Therefore, the length of time an individual spends in the exposed state is distributed according to an exponential distribution with parameter $\gamma$. The same reasoning, but using assumption (4) instead, shows that the length of time spent infectious is exponentially distributed with parameter $\delta$. The mean of an exponential distribution with parameter $\alpha$ is $1/\alpha$. Therefore, the mean length of time spent by an individual in states $E$ and $I$ is $1/\gamma$ and $1/\delta$ respectively.

### 4.3.1 Basic reproductive number $R_0$ for well-mixed transmission models

In the well-mixed model, $R_0$ is defined as the expected number of secondary cases transmitted directly by a single infectious individual in an otherwise susceptible population. As discussed beforehand, the mean length of time an individual spends infectious is $1/\delta$. The infection is assumed to spread to $\beta S(0)$ individuals per unit time. Therefore, the value of $R_0$ is given by Eq. (4.2).

$$R_0 = \frac{\beta S(0)}{\delta} \tag{4.2}$$

### 4.3.2 Final size relation

Let $R_\infty$ denote the final size of the epidemic, the number of individuals who, once the epidemic has ended, were infected at any point during the epidemic. We wish to express $R_\infty$ as a function of $R_0$. Dividing $\frac{dR}{dt}$ by $\frac{dS}{dt}$ gives $\frac{dR}{dS} = -\frac{\delta}{\beta S}$. Integration of both sides with respect to $S$, and evaluation at 0 to find the integration constant ($R(0) = 0$), gives:

$$S(t) = S(0)\exp\left(-\frac{\beta R(t)}{\delta}\right),$$

with $S(0) = N - 1$, as the population is otherwise susceptible apart from a singular infectious individual. We take the limit at $t \to \infty$ and let $S_\infty = S(\infty)$ and $R_\infty = R(\infty)$. Dividing both sides by $N$, and using the relationship $R_0 = \frac{\beta S(0)}{\delta}$ in Eq. (4.2), we have:

$$\frac{S_\infty}{N} = 1 - \frac{R_\infty}{N} = \frac{S(0)}{N}\exp\left(-\frac{R_0 N}{S(0)}\frac{R_\infty}{N}\right).$$

Using Wang (2010), we have that the $R_\infty$ as a function of $R_0$ is given by Eq. (4.3).

$$\frac{R_\infty}{N} = 1 + \frac{S(0)}{R_0 N} W\left(-R_0 \exp\left(-\frac{R_0 N}{S(0)}\right)\right) \tag{4.3}$$

Here $W$ denotes the Lambert $W$ function. As $N \to \infty$, the ratio $\frac{S(0)}{N} \to 1$, and Eq. (4.3) becomes:

$$\frac{R_\infty}{N} = 1 + \frac{1}{R_0} W\left(-R_0 \exp\left(-R_0\right)\right) \tag{4.4}$$

Fig. 4.3 shows a plot of the relationship given in Eq. (4.3) for three finite populations of size $N$ (40, 160, and 1000), in addition to the relationship given in Eq. (4.4) for an infinitely large population. As the population grows, the relationship tends to that of the infinitely large population.

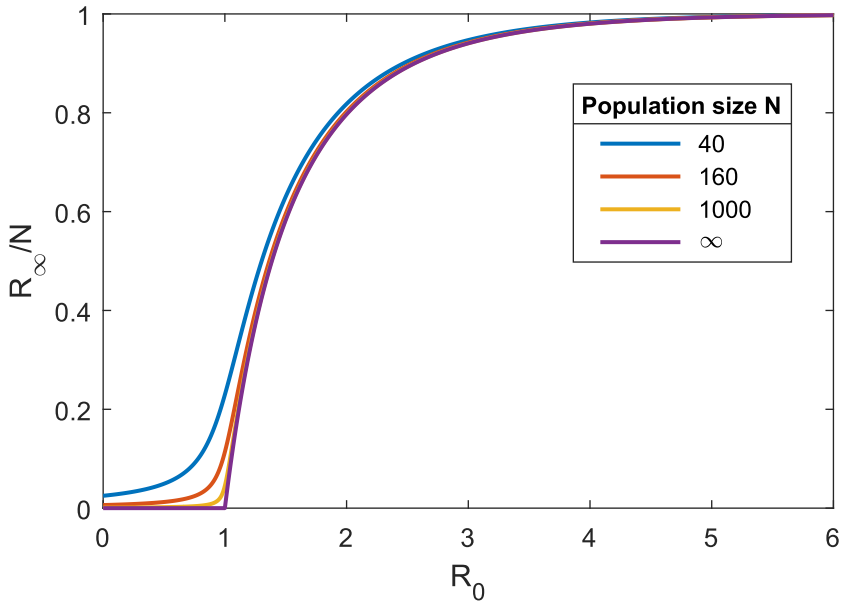



Figure 4.3: **Plot of the relationship between basic reproductive number $R_0$ and final epidemic size $R_\infty$, in an SEIR-model performed on populations of size $N$.** As $N \to \infty$, the relationship shown in the finite populations tends to that of the infinitely large population.

## 4.4 Stochastic Transmission Model

A well-mixed model is not appropriate when studying small populations, as assumptions, including uniform contact structure and contact frequency, can no longer be assumed [Britton, 2010]. As such, we employ weighted contact networks to incorporate this contact heterogeneity.

Our stochastic transmission model has two components: the underlying weighted population network, and the epidemic simulation model. Our population is assumed, as with the classical SEIR-model, to be of size $N$ and the population size remains constant because the rates of birth and non-disease-induced mortality are zero. A contact occurs between two individuals if they come within 1 metre of each other. This value reflects the detection distance of the collars used on the real-world possum population in Section 2.2. Since the population is assumed to not be well-mixed, not every pair of individuals will come into contact.

The underlying structure of our population is modelled using a weighted contact network $G$. Each individual $i$ is represented by a node in $G$. Edges in $G$ connect nodes, and represent the possibility of a contact occurring between pairs of individuals. In order to account for the heterogeneity of individuals' contacts, a nonzero value, the edge weight is associated to each edge. The weight of the edge between $i$ and $j$ is denoted $w_{ij}$, which is the contact rate for the pair.

We use an SEIR compartmental model to simulate an epidemic spreading through our population. The state of the population at time $t$ is given by $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$ where, for individual $i$, $x_i(t)$ is assigned a label $S$, $E$, $I$ or $R$, depending on its current status (we write $x_j(t) = S$ to denote that at time $t$, individual $j$ is susceptible). We denote $\mathbf{x}(0) = (x_1(0), \ldots, x_N(0))$ to be the initial state of the population.

The rate at which contacts between a susceptible individual $i$ and an infectious individual $j$ result in the disease being passed from $j$ to $i$ is given by $\beta$. Thus, the transition rate from the susceptible to the exposed state for individual $i$ at time $t$, given by $\kappa_i(t)$, is equal to $\beta$ multiplied by sum of the edge weights of infectious individuals incident to $i$. Exposed individuals become infectious at a constant rate $\gamma$. Infected individuals die, or are removed, from the population at a constant rate $\delta$. These assumptions correspond to individuals remaining in compartments $E$ and $I$ for exponentially distributed times (with means $1/\gamma$ and $1/\delta$ respectively), which is consistent with the classical SEIR-model. When an individual is removed, no further contacts with it

can occur, thus it cannot transmit the infection to any other individuals. The transition rates are summarised in Table 4.1. We assume that $\beta$, $\gamma$ and $\delta$ are identical for each individual.

| Current state $x_i(t)$ | Transition | Transition rate $q_i(t)$ |
|---|---|---|
| $S$ | Susceptible to Exposed ($S \to E$) | $\kappa_i(t) = \beta \sum_{x_j(t)=I, j \neq i} w_{ij}$ |
| $E$ | Exposed to Infectious ($E \to I$) | $\gamma$ |
| $I$ | Infectious to Removed ($I \to R$) | $\delta$ |

Table 4.1: We calculate the rates of transition $q_i(t)$, between the four different states, that an individual $i$ in our population can undergo at time $t$.

At time 0, a small proportion of the individuals, referred to as the *seed population*, have been exposed to the disease, and thus $x_i(0) = E$ for those individuals. The individuals in the seed populations are referred to as *seed cases*. The remainder are susceptible. Any infectious individual who is not in the seed population is referred to as a *secondary infection*, and an epidemic with at least one secondary infection is defined to be an *established epidemic*. We implement a Gillespie algorithm, as described in Section 4.2.3, to simulate the bovine tuberculosis epidemic. At each time $t$, the total transition rate is given by Eq. (4.5).

$$q(t) = \left( \sum_{s \in S(t)} \kappa_s(t) \right) + \gamma E(t) + \delta I(t) \tag{4.5}$$

## 4.5 Basic Reproductive Number $R_0$ for network models

As discussed in Section 4.2.2, the basic reproductive number $R_0$ is defined as the expected number of secondary infections directly transmitted from a singular infectious individual in an otherwise susceptible population. In Section 4.3.1, we derived a formula for $R_0$ in the case of a well-mixed transmission model performed on a population of size $N$, showing that it was equal to the product of the number of individuals infected per unit time ($\beta S(0)$) and the average time spent infectious ($1/\delta$). The well-mixed model assumes that each pair of individuals come into contact, and do so at the same rate. This assumption cannot be applied to our stochastic transmission model, as both the contact structure and rate of contact are not homogeneous across the population. In this section, we review the previous work conducted to find an analytical formula for $R_0$ in epidemic network models.

Building directly from the well-mixed model, the intuitive formula for $R_0$ is given by Eq. (4.6), where $\langle k \rangle$ is the mean degree of the network's adjacency matrix. The reasoning behind this is that for an initially infectious node in a network of otherwise susceptible nodes, the expected number of secondary infections is proportional to the average number of connections of the infectious node, that is $\langle k \rangle$. This formula has been used when modelling epidemics on both homogeneous networks (e.g. complete graphs), and Erdős-Rényi-Gilbert random networks [Kephart and White, 1992]. However, when the contact structure of the network is sufficiently heterogeneous, the assumptions that allow the use of this formula are not applicable. For example, the as the degree distribution of a network becomes more broad, the mean degree may no longer be an accurate way of modelling the rate of new infections by the initially infectious node.

$$R_0 = \frac{\beta \langle k \rangle}{\delta} \tag{4.6}$$

For networks with more heterogeneity, a formula for $R_0$ was developed that reflects the effect of the broader degree distribution on the epidemic spread. The formula, given in Eq. (4.7), is proportional to the ratio of the first and second moments of the degree distribution. First used to better understand HIV transmission dynamics, this formula was developed to reflect the effect that highly connected nodes in the network have on HIV epidemic spread [Anderson et al., 1986, May and Anderson, 1987, May and Anderson, 1988]. This formula has been applied to not only the real-world sexual contact networks in Anderson et al. (1986) and subsequent papers, but also scale-free networks (in particular certain Barabási-Albert networks) [Pastor-Satorras and

Vespignani, 2002] and uncorrelated random networks [Boguná and Pastor-Satorras, 2002].

$$R_0 = \frac{\beta\langle k^2\rangle}{\delta\langle k\rangle} \tag{4.7}$$

A further development was made that related the basic reproductive number $R_0$ to the eigenvalues of the adjacency matrix. Wang et al. (2003) proposed a formula that did not rely on any particular network topology. From their SIS-model, where individuals can move to and from Susceptible and infectious states, they concluded that, for a given network with adjacency matrix $A$, $R_0$ is proportional to the spectral radius, the largest real eigenvalue, of $A$. This formula is given in Eq. (4.8). In the same paper, they showed that the formula, when applied to homogeneous networks or Erdős-Rényi-Gilbert random networks, Eq. (4.8) is equivalent to Eq. (4.6). Further, it was also shown that for finite scale-free networks, Eq. (4.8) is a more precise indicator for the epidemic threshold than Eq. (4.8).

$$R_0 = \frac{\beta\rho(A)}{\delta}, \text{ where } \rho(A) \text{ is the spectral radius of } A \tag{4.8}$$

However, Van Mieghem et al. (2008) concluded that model of Wang et al. (2003) was only accurate when $\beta/\delta$ was below the steady-state epidemic threshold. Furthermore, Omic (2010) discussed a potentially erroneous assumption that Wang et al. (2003) made, that there is a probability of 0.5 that node $i$ becomes cured after being infected by its neighbours, noting that the assumption of the 0.5 probability was not an valid modelling step. The model in Wang et al. (2003) was later improved upon in Chakrabarti et al. (2008) [Van Mieghem, 2011, Cator and Van Mieghem, 2012]. Youssef and Scoglio (2011) showed that this epidemic threshold holds for SIR-epidemics taking place on networks. The networks we discuss throughout this thesis are weighted random geometric graphs, and therefore of particular relevance is the application of this formula by Preciado and Jadbabaie, (2009) and Estrada et al. (2016) to random geometric graphs or their extensions.

## 4.6 Outputs from model simulations

We have now presented our stochastic transmission model, which allows us to simulate an epidemic spreading from an initial group of exposed individuals, through a population represented by a weighted contact network. In this section we detail the outputs that can be produced from a single epidemic simulation, and how results from multiple simulations can be consolidated. The final part, concerning consolidating results, will be of importance in Chapter 5, where we we perform multiple epidemic simulations on networks produced using the models described in Chapter 3.

### 4.6.1 Outputs from a single model simulation

From each simulation of an epidemic, we can produce several results:

- A time series of the epidemic. At the end of each 24-hour period, we count the number of individuals in each state.
- The final size $R_\infty$ of the epidemic, the number of individuals who have been removed from the population during the simulation. Since an individual can only be removed if they were infectious, this is equivalent to the number of individuals who were in state $I$ at any point throughout the simulation.
- The directed 'infection trees' of the epidemic. For each infected individual $i$, our model gives, explicitly, the individual $j$ who infected $i$. We can create disjoint directed trees, each representing the infection spreading from a single individual in the seed population.
- The number of infections transmitted directly by a single initially infected node, a seed case, in a otherwise susceptible population, during the period in which the node is infectious.
- The numbers of days, from 0, until the last infection and last removal events.

To illustrate this, an example binary random geometric graph, with network density 0.2, is produced by placing 40 vertices in a circle of radius 1 metre according to a spatial Poisson point process. Pairwise distances between vertices are calculated, and edges representing the 156 pairs (thereby producing a network density of 0.2) with the shortest pairwise distances are chosen. This example network is shown in Fig. 4.4.

We perform a single simulation of an epidemic spreading on this network. In our simulation, we set $\beta = 0.01$, corresponding to 1% of contacts causing infection. Further we set $\gamma = 0.0166$ and $\delta = 0.0133$, which correspond

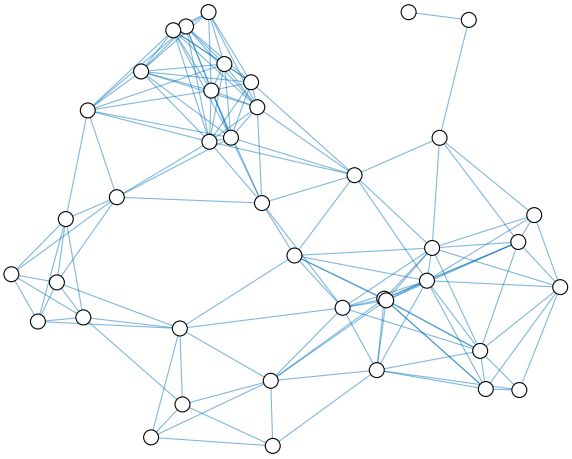


Figure 4.4: **An example random geometric graph on 40 vertices, representing a population of 40 individuals.** This network has network density 0.2, by the explicit choice of 156 edges corresponding to the 156 shortest distances between individuals.

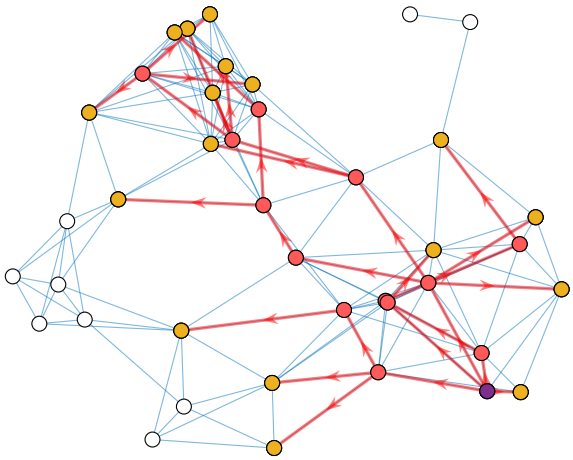


Figure 4.5: **The 'infection tree' of an epidemic simulation.** From an epidemic simulation, we produce the directed tree that represents the spread of disease through the population. The infection tree for the simulation also shown in Fig. 4.6 is highlighted in red, with the seed case highlighted in purple, and infected individual who did not infect anyone themselves in yellow.

to latent and infectious periods of 60 days and 75 days respectively. The epidemic started with a single seed case, chosen from the population at random, and was run for 4000 days. The infection tree of the epidemic is shown in Fig. 4.5. The seed case is highlighted in purple, and infected individual who did not infect anyone themselves are highlighted in yellow.

The values of $S(t)$, $E(t)$, $I(t)$, and $R(t)$ at the end of each 24-hour period following the start of the epidemic are calculated. Fig. 4.6 shows the plot of these values. The final size of this epidemic was 31. There were five individuals to whom the infection was directly transmitted by the seed case. The last infection event occurred on day 562, and the last removal event on day 750.

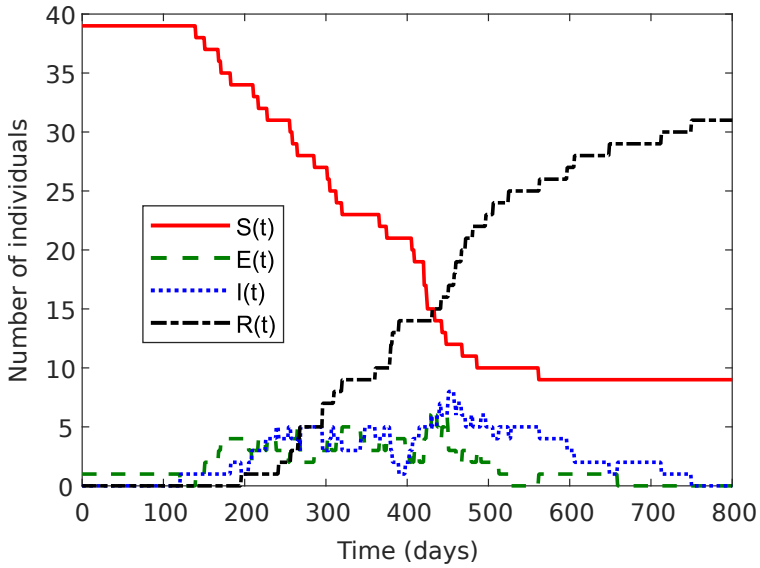



Figure 4.6: **Plot of values of $S(t)$, $E(t)$, $I(t)$, and $R(t)$.** A single simulation of an epidemic spreading on the network shown in Fig. 4.4 is performed. The numbers of individuals in each epidemic compartment at the end of each 24-hour period is calculated.

### 4.6.2 Consolidating the results of multiple epidemic simulations

When performing multiple simulations, we consolidate the results in the following ways:

- Calculating the average values of $S(t)$, $E(t)$, $I(t)$, and $R(t)$ for each 24-hour period.
- Producing distributions of:
  - The final size $R_\infty$ of each epidemic.
  - The number of infections directly transmitted to an individual by a seed case. By taking the average value of this distribution, we calculate an estimation for the basic reproductive number $R_0$.
  - The days on which the final infection and final removal events took place.

To illustrate this, we perform ten simulations with the same population, parameters and initial conditions as Section 4.6.1. Fig. 4.7 shows the distribution of the compartmental values for each simulation (in green), and the average value over the ten simulations (in red). For simulations where the epidemic established itself, the compartment numbers follow similar trends, with the curves shifted due the stochastic nature of the seed case's latency period.

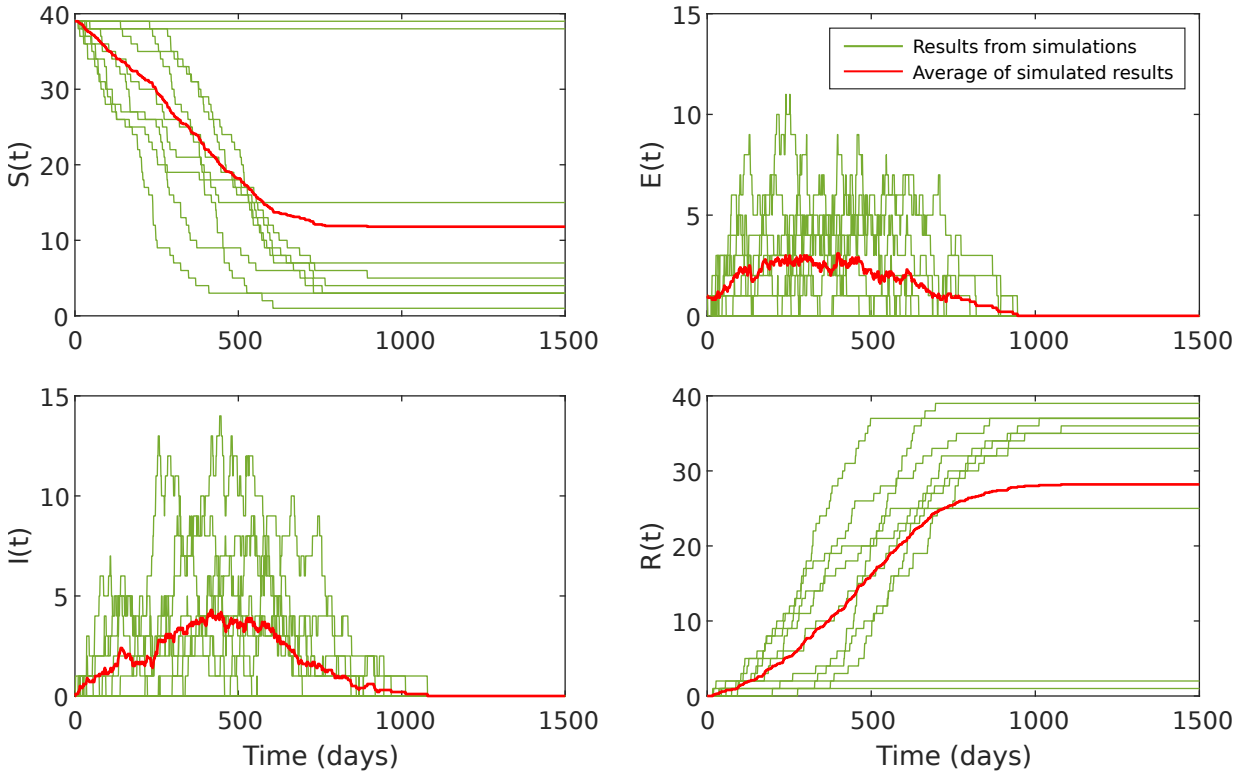



Figure 4.7: **Numbers of individuals in each compartment, for each 24-hour period, from ten epidemic simulations.** The same population, parameters and seed case as with Section 4.6.1 are used. Plots of $S(t)$, $E(t)$, $I(t)$, and $R(t)$ are shown in green, with the average values in red.

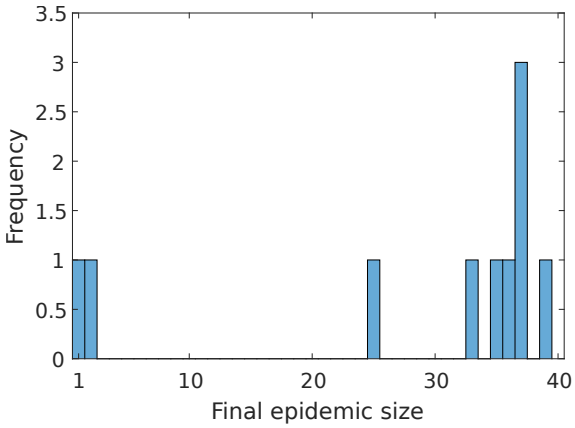



Figure 4.8: **Distribution of final epidemic sizes $R_\infty$.** There is one simulation where the epidemic failed to establish, and thus $R_\infty = 1$, and another where only one individual was removed. The majority of simulations resulted in a large proportion of individuals in the population being removed.

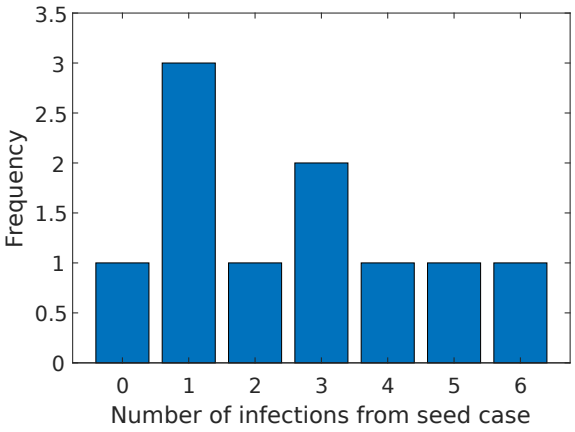



Figure 4.9: **Distribution of the number of infections directly transmitted by the seed case.** The modal number of infections from this single individual is one. By taking the average over the ten simulations, we produce an estimated value of $R_0 = 2.6$.

The distribution of final epidemic sizes $R_\infty$, the number of individuals who were infectious during the epidemic, is shown in Fig. 4.8. There is one simulation where the epidemic failed to establish, and another where only one individuals was infected. The majority of simulations resulted in a large proportion of individuals in the population being infected.

From each simulation, the number of infections directly transmitted by the single individual in the seed population is calculated. The distribution of these values is given in Fig. 4.9. The modal number of infections from this single individual is one. By taking the average over the ten simulations, we produce an estimated value of $R_0 = 2.6$. Fig. 4.10 shows the distributions of final exposure and final removal event times.

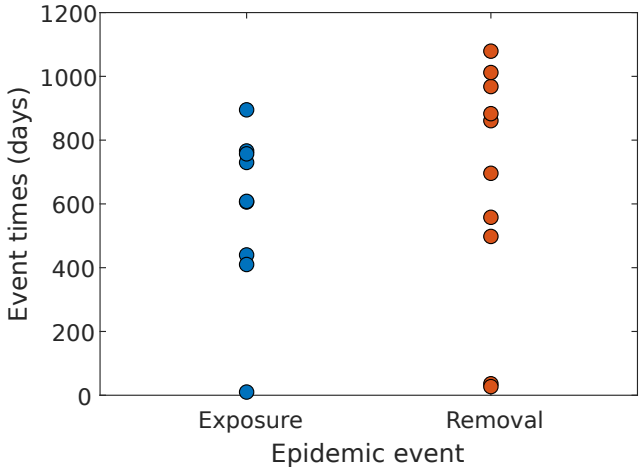



Figure 4.10: **Distributions of the final exposure and final removal event times.**

### 4.6.3 Comparison of analytical and simulated values of $R_0$

As discussed in Section 4.5, there are various formulae to calculating the value of $R_0$ for an epidemic on a network. We can also calculate a value of $R_0$ from epidemic simulations, as explained in Section 4.6.2. Fig. 4.11 shows a comparison between simulated values of $R_0$ and values calculated using Eqs. (4.6) to (4.8), for three different classes of network: Erdos-Reyni-Gilbert, small-world, and random geometric graphs. Regardless of which analytical formula is used, Fig. 4.11 shows that the simulated value is always lower than the analytical value. This could be because we have in fact chosen an erroneous definition for our simulated $R_0$ value.

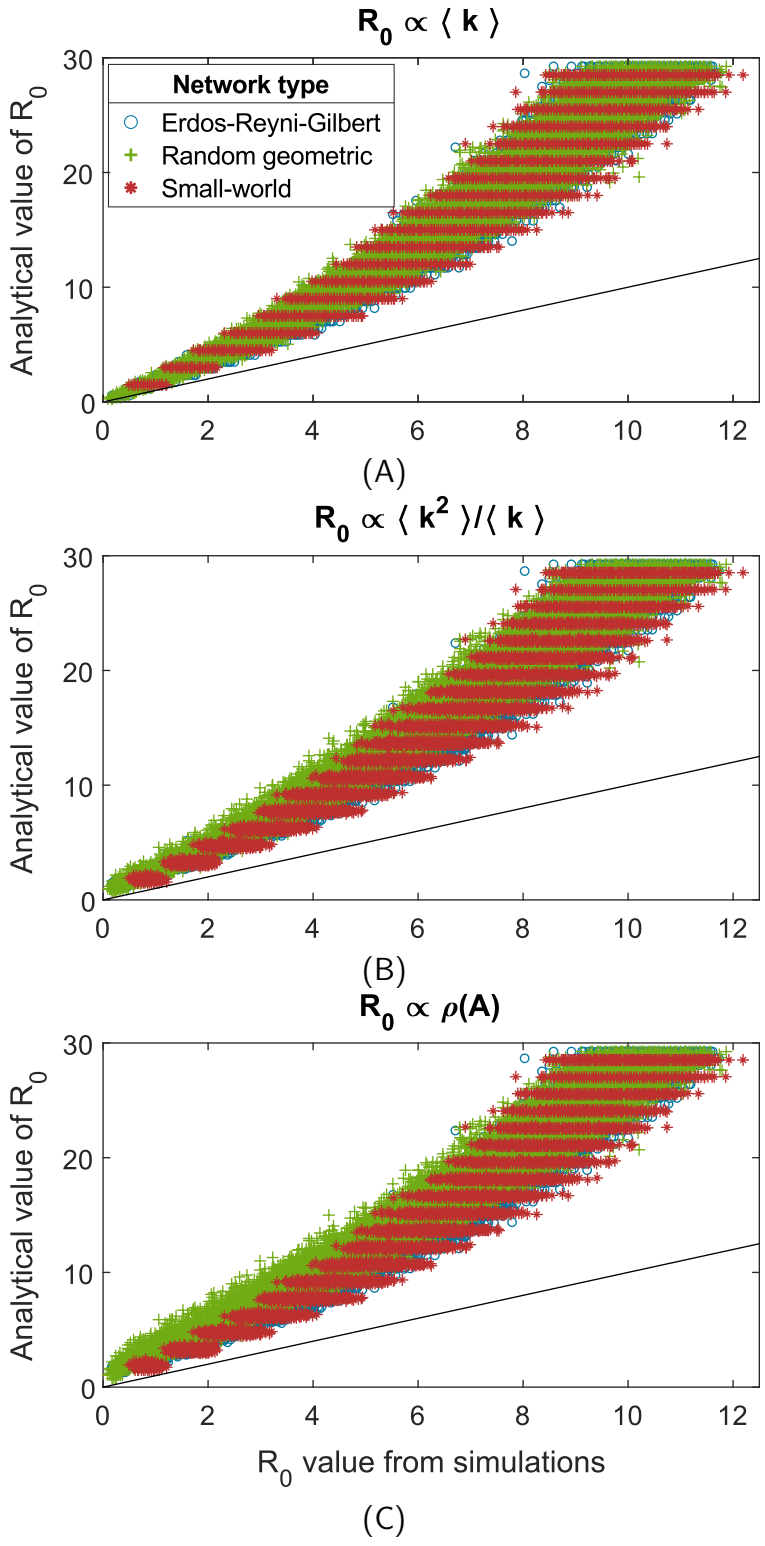



Figure 4.11: **Plot of simulated against analytical values of $R_0$, for three classes of network.** For Erdos-Reyni-Gilbert, small-world, and random geometric networks, we calculate analytical values of $R_0$ using Eqs. (4.6) to (4.8). Further, we perform epidemic simulations, calculating simulated values of $R_0$ as per Section 4.6.2. The black line is the line $y = x$. Regardless of which analytical formula is used, the simulated value is always lower than the analytical value. This could be because we have in fact chosen an erroneous definition for our simulated $R_0$ value.

## 4.7 Effect of changing parameters and network topology on an epidemic

In this section, we aim to better understand how different epidemic parameters and topological properties of networks affect the spread of an epidemic. To achieve this, simplified toy networks are produced, and SEIR-model epidemics are simulated on them. For each change of parameter, we produce 50 binary networks on 40 vertices, in order to reduce the effect of a single network's topology on the results. Each population is produces using the same network construction method as in the previous subsection. For each network, 1000 epidemics were simulated, with different seed cases. Seed cases were chosen randomly from a possible seed population, where the possible seed population is determined by a rule that we will explicitly state. We investigate the effect of changing:

- The values of the epidemic parameters $\beta$, $\gamma$, and $\delta$.
- The network density.
- The process by which the possible seed population is determined.
- The edge weight distribution.

Throughout this section, we will be comparing results from the stochastic transmission model discussed in Section 4.4, with those of the well-mixed model, examining and highlighting any parameters, network structures, or edge weight distributions makes the stochastic model to converge to the well-mixed model.

As an initial comparison, we compare the results from the stochastic transmission model with both a deterministic and a stochastic version of the well-mixed model. To model a stochastic version of the well-mixed model is described in Section 4.3, we implement a Gillespie algorithm, as defined in Section 4.2.3. At each time $t$, the total transition rate is given by $q(t) = \beta S(t)I(t) + \gamma E(t) + \delta I(t)$. We perform multiple simulations of the stochastic well-mixed model, in addition to our stochastic transmission model, performed on a complete network, and summarise the results as discussed in Section 4.6.2. As shown in Fig. 4.12, the results from the stochastic well-mixed model match closely with those of the stochastic transmission model. This is to be expected, as simulating the model on a complete network is equivalent to the well-mixed assumption.

However, the results from the deterministic well-mixed model do not match up with those of either the stochastic well-mixed or stochastic transmission models. This is most likely due to stochasticity; there is a

non-zero probability that the epidemic will die out before infecting the entire population, as it does in the deterministic model. Further, the stochasticity of individuals transitioning from the Exposed to the Infectious compartments will also result in a difference between the stochastic and deterministic well-mixed models.

As the stochastic well-mixed model more closely matches our stochastic transmission model, in this section we will use results from the stochastic well-mixed model when comparing with our transmission model.

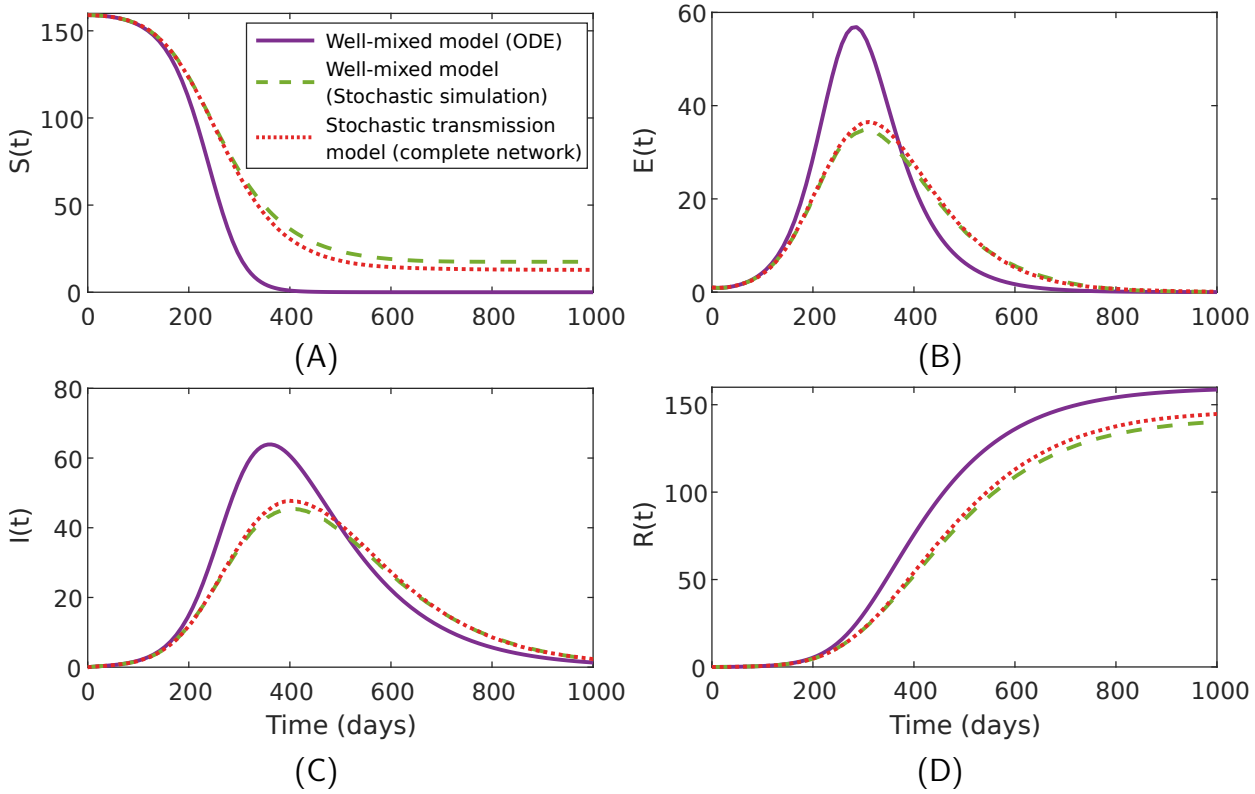



Figure 4.12: **Comparison between results from the deterministic and stochastic well-mixed models, in addition to the stochastic transmission model.** The deterministic well-mixed model does not match up with either the stochastic well-mixed or stochastic transmission models. This is most likely due to stochasticity; there is a non-zero probability that the epidemic will die out before infecting the entire population as it does in the deterministic model. Further, the stochasticity of individuals transitioning from the Exposed to the Infectious compartments will also result in a difference between the stochastic and deterministic well-mixed models. The stochastic well-mixed model, however, does match more closely matches our stochastic transmission model.

### 4.7.1 Altering the values of the epidemic parameters $\beta$, $\gamma$, and $\delta$

In order to assess the effect of changing one of the transmission coefficient $\beta$, rate of becoming infectious $\gamma$, and rate of removal $\delta$, network density is fixed at 0.2, each individual in each population is a possible seed case, and initial values of $\beta$, $\gamma$, and $\delta$ are set. When a single parameter is changed, the values of the other parameters remain the same. Three values of each parameter were tested. The initial and tested values are given in Table 4.2.

| | $\beta$ | $\gamma$ | $\delta$ |
|---|---|---|---|
| Initial value | 0.01 | $\frac{1}{60}$ | $\frac{1}{75}$ |
| Tested values | 0.005, 0.01, 0.1 | $\frac{1}{30}$, $\frac{1}{60}$, $\frac{1}{120}$ | $\frac{1}{30}$, $\frac{1}{75}$, $\frac{1}{150}$ |

Table 4.2: The initial and tested values for the the transmission coefficient $\beta$, rate of becoming infectious $\gamma$, and rate of removal $\delta$.

When looking at the average values of $S(t)$, $E(t)$, $I(t)$, and $R(t)$ for each 24-hour period, as shown in Figs. 4.13 to 4.15, the following conclusions can be drawn:

- Increasing the value of the transmission coefficient $\beta$ results in a higher average value of the final epidemic size $R_\infty$. It also results in more infection events in a shorter period of time, as shown by the values of $E(t)$, but the epidemic itself ends, on average, much more quickly.

- Increasing the rate of becoming infectious $\gamma$, and thereby decreasing the length of the latent period, has no effect on the final epidemic size. This is to be expected. A shorter latent period results in individuals becoming infectious quicker, but the the values of $\beta$ and $\delta$ remaining constant means that the same number of infection events will occur, only in a shorter period of time, causing the epidemic to end earlier.

- Decreasing the length of the infectious period, that is increasing the removal rate $\delta$, results in fewer infections, and the epidemic dying slightly sooner. This is due to the shorter infectious period leading to fewer infection events before an individual is removed from the population.

These conclusions are corroborated by the distributions of final infection and final removal events shown in Figs. 4.16 and 4.17.

Similar conclusions can be drawn when looking at the distributions of the final epidemic sizes, as shown in Fig. 4.18. As the transmission coefficient $\beta$ increases, or the removal rate $\delta$ decreases, the number of established epidemics (involving at least one infection from the seed population) increases. Further, as $\beta$

increases, the overall final size of the epidemic increases as well. In fact, once $\beta \geq 0.1$, most simulations result in the epidemic infecting almost all individuals in a population. This results from a larger $\beta$ allowing for more infections before the individuals are removed from the population. A similar conclusion, with similar justification, can be made when $\delta \leq 1/150$.

As shown in Fig. 4.19, altering the value of $\gamma$, the rate of becoming infectious, does little to change the distribution of $R_0$, though the mean value does increase as $\gamma$ is decreased. This is possibly due to the time spent in the exposed compartment. If an individual spends less time in the latent state, then they may infect individuals who would have been infected by the seed case if the latent period were longer. This would then result in a smaller value of $R_0$. Increasing the value of $\beta$, or decreasing the value of $\delta$, results in distributions with higher values, and higher means.

When comparing the results between our stochastic transmission model and the well-mixed model, the only parameter change that allows the two methods to converge is by increasing the value of the transmission parameter $\beta$. When we increase $\beta$ to 0.1, the results, shown in Figs. 4.13 and 4.16 to 4.18, match closely with those of the well-mixed model. This is mostly likely due to the higher value of $\beta$ causing a similar number of infections to occur, allowing the epidemic to spread in a similar way to the well-mixed model, even with the lower network density prohibiting all transmission routes fro mthe well-mixed model to be available to the epidemic. However, the same phenomenon is not seen when the values of $\gamma$ or $\delta$ are changed. This could be due to the fixed value of $\beta = 0.01$ used, or that the values of $\gamma$ and $\delta$ are not sufficient for the epidemic to progress in the manner it does in the well-mixed model.

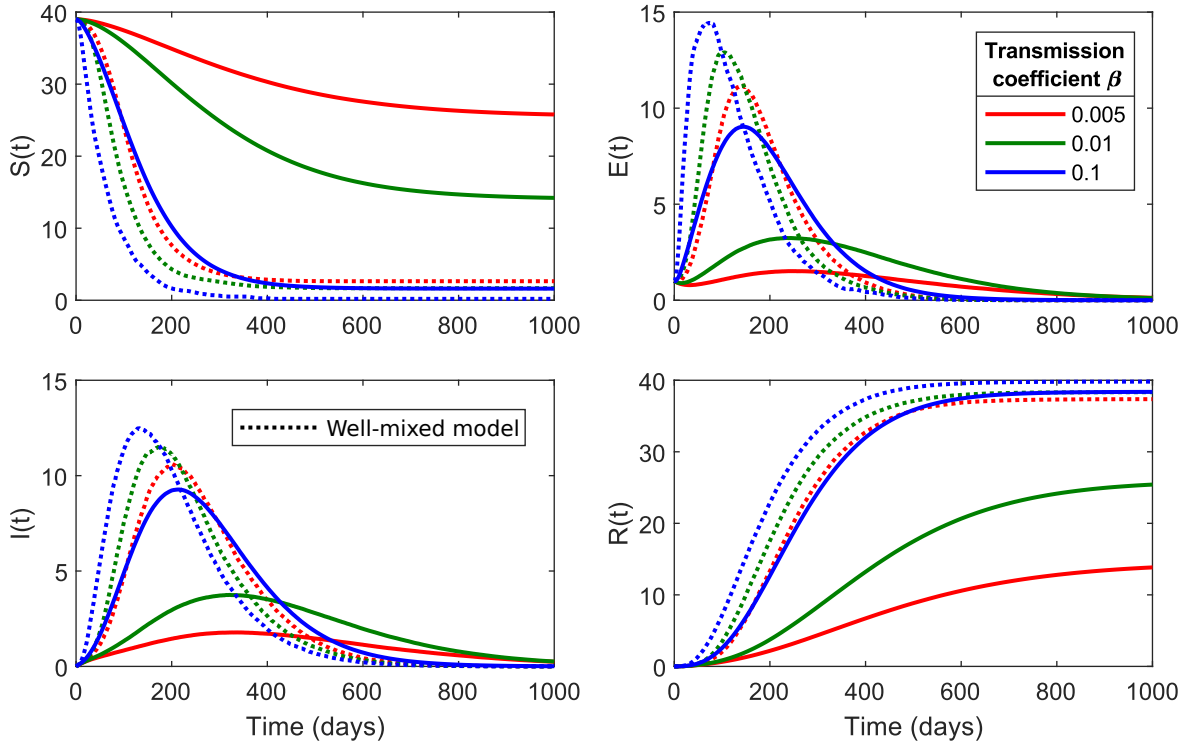



Figure 4.13: **Average numbers of individuals in each compartment, for each 24-hour period, for three different values of the transmission coefficient $\beta$.** As the value of $\beta$ increases, the average value of the final epidemic size $R_\infty$ increases. Additionally, more infection events occur in a shorter period of time, as shown by the values of $E(t)$, but the epidemic itself ends, on average, more quickly.

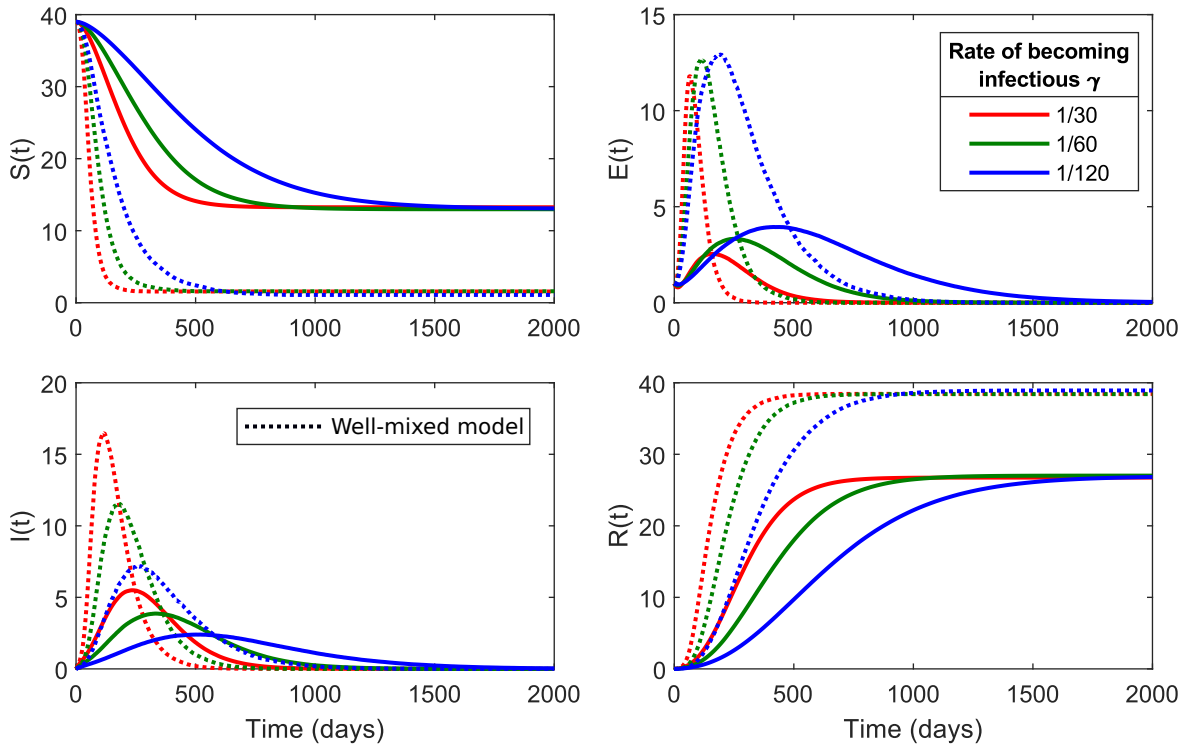



Figure 4.14: **Average numbers of individuals in each compartment, for each 24-hour period, for three different values of $\gamma$, the rate of becoming infectious.** By increasing the rate of becoming infectious $\gamma$, and thereby decreasing the length of the latent period, individuals become infectious quicker, but as the the values of the transmission coefficients $\beta$ and the removal rate $\delta$ remain constant, the same number of infection events will occur, only in a shorter period of time. This results in an epidemic ending earlier, but also little to no effect on the final epidemic size $R_\infty$.

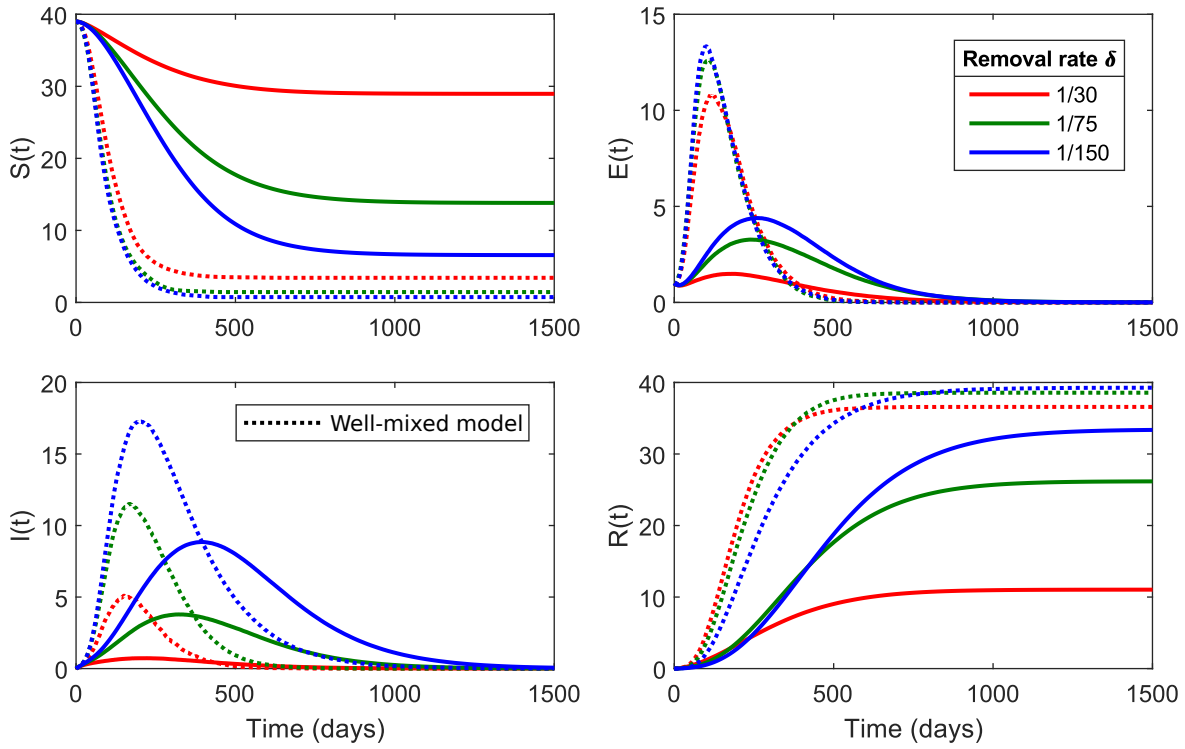



Figure 4.15: **Average numbers of individuals in each compartment, for each 24-hour period, for three different values of $\delta$, the rate of removal.** Increasing the removal rate $\delta$, that is decreasing the length of the infectious period, results in fewer infections, and thus both a smaller value of the final epidemic size $R_\infty$ and the epidemic dying slightly sooner. This is due to the shorter infectious period leading to fewer infection events before an individual is removed from the population.

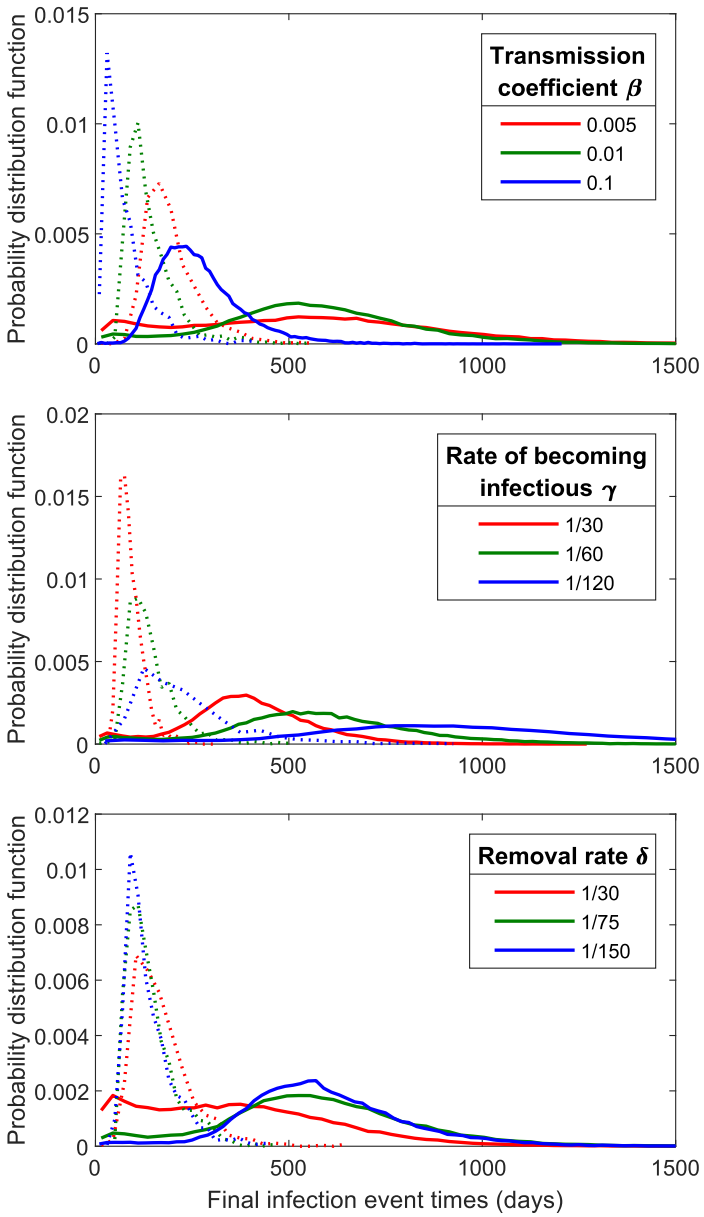



Figure 4.16: **Plots of the final infection events for each epidemic simulation.** Increasing either of the transmission coefficient $\beta$ or the rate of becoming infectious $\gamma$, causes the final infection events to occur sooner. This is to be expected as a higher value of either parameter causes infection events to occur more rapidly ($\beta$) or exposed individuals to become infectious sooner ($\gamma$). Increasing the removal rate $\delta$ does decrease the final event time, but the effect is less than altering $\beta$ or $\gamma$.

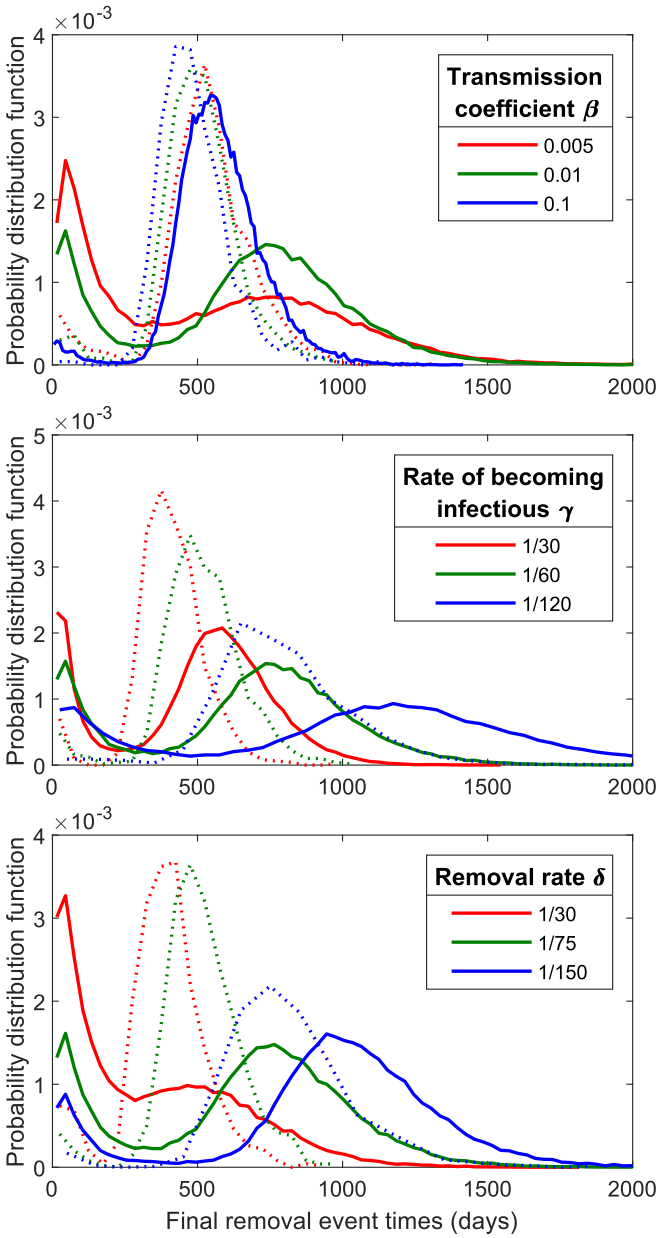



Figure 4.17: **Plots of the final removal events for each epidemic simulation.** Increasing the removal rate $\delta$, that is decreasing the length of the infectious period, results in fewer infections, and thus both a smaller value of the final epidemic size $R_\infty$ and the epidemic dying slightly sooner. This is due to the shorter infectious period leading to fewer infection events before an individual is removed from the population.

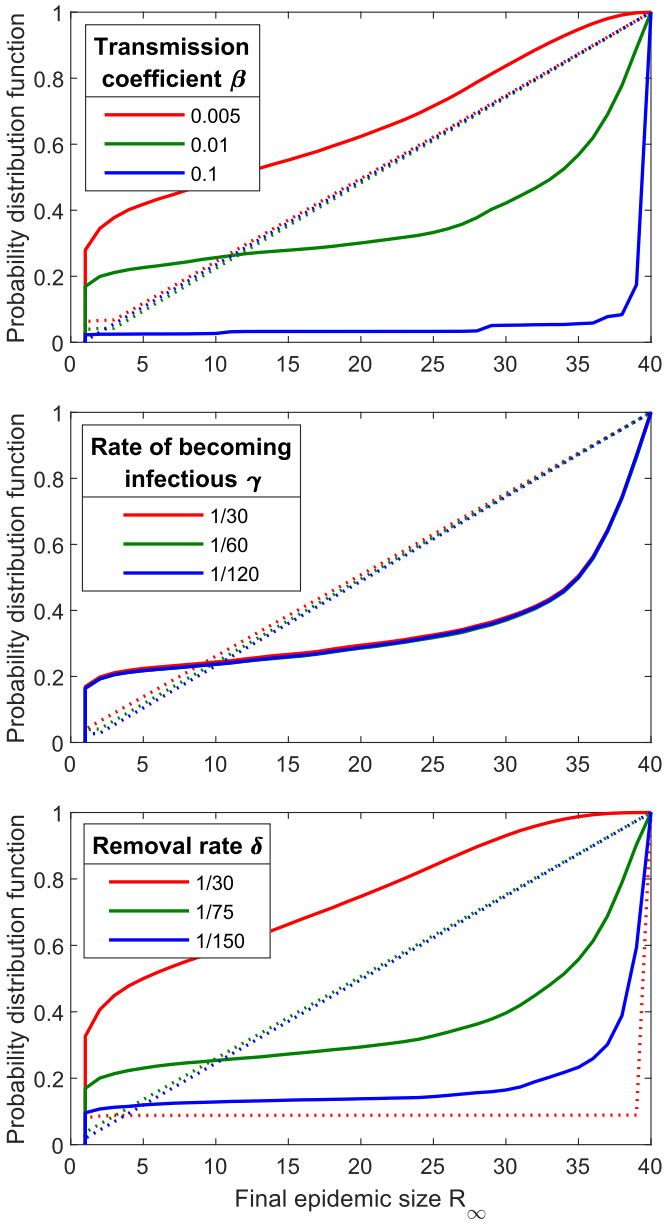



Figure 4.18: **Cumulative distribution functions of the final epidemic sizes.** Similar conclusions can be drawn when looking at the distributions of the final epidemic sizes, as shown in Fig. 4.18. As the transmission coefficient $\beta$ increases, or the removal rate $\delta$ decreases, the number of established epidemics (involving at least one infection from the seed population) increases. Further, as $\beta$ increases, the overall final size of the epidemic increases as well. In fact, once $\beta \geq 0.1$, most simulations result in the epidemic infecting almost all individuals in a population. This results from a larger $\beta$ allowing for more infections before the individuals are removed from the population. A similar conclusion, with similar justification, can be made when $\delta \leq 1/150$. Results from the well-mixed model are shown as dashed lines, for comparison.

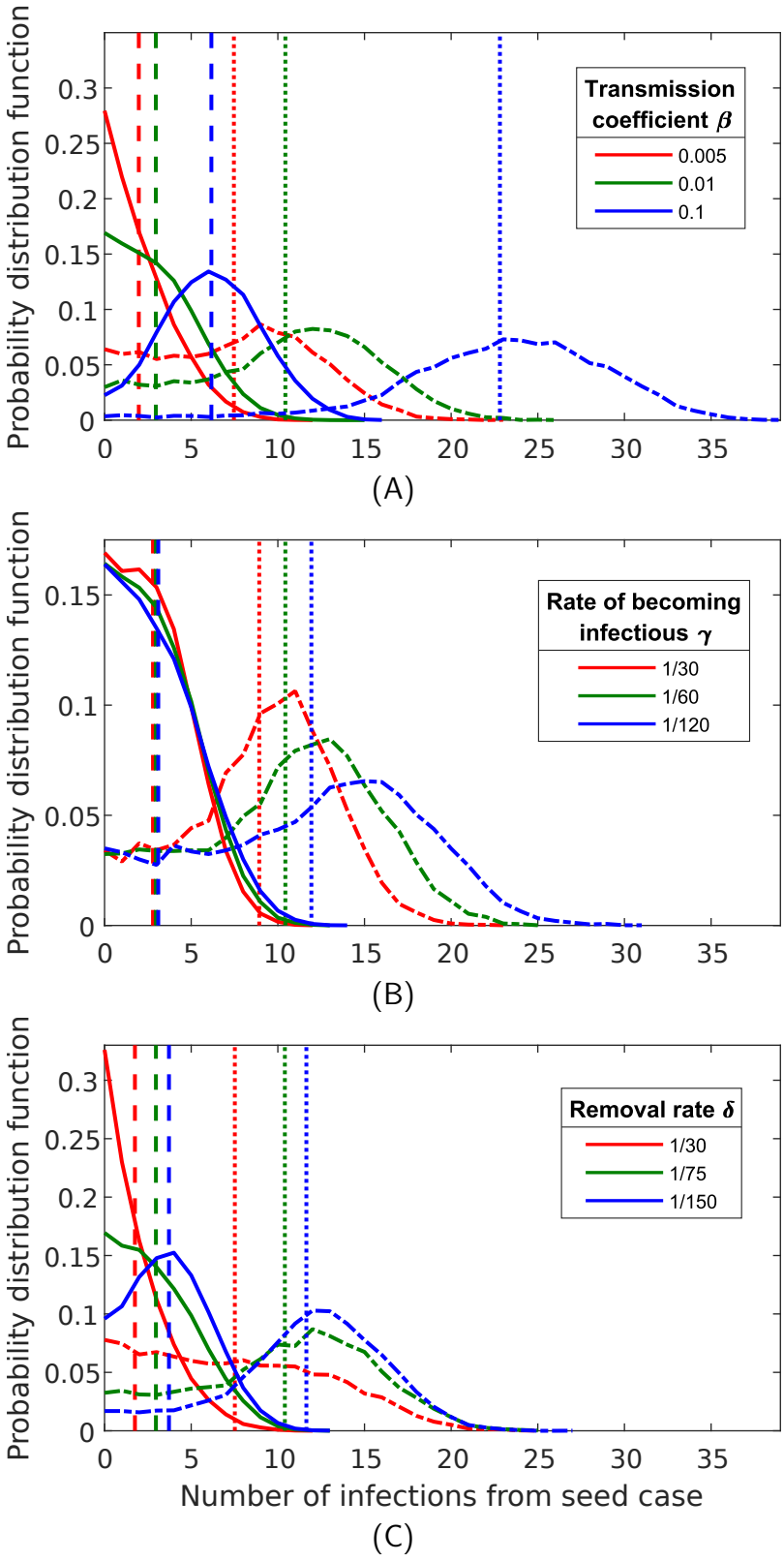



Figure 4.19: **Distributions of the number of infections directly transmitted from the seed case.** Increasing the value of the transmission coefficient $\beta$, or decreasing the value of the removal rate $\delta$, results in more individuals being infected by the seed case. However, altering the value of the rate of becoming infectious $\gamma$ does little to change the distribution. The mean values of these distributions, namely values of $R_0$, are shown as dash-dot lines. As $\beta$ is increased, or $\delta$ is decreased, the value of $R_0$ is increased. The value of $R_0$ does increase as $\gamma$ is decreased. This is possibly due to the time spent in the exposed compartment. If an individual spends less time in the latent state, then they may possibly infected individuals who, if the latent period were longer, would be infected by the seed case. This would then result in a smaller value of $R_0$. Results from the well-mixed model are shown as dashed lines, for comparison.

### 4.7.2 Changing the network density

Network density is the proportion of edges that exist in a network, out of the total that can possibly exist. Changing the network density affects the spread of an epidemic through a network as a higher network density corresponds to more edges in the networks, allowing the epidemic more opportunity to spread. We start by fixing the transmission coefficient $\beta$, rate of becoming infectious $\gamma$, and removal rate $\delta$, at the following values: $\beta = 0.01$, $\gamma = \frac{1}{60}$, and $\delta = \frac{1}{75}$. Each individual in each population is a possible seed case. Ten values of the network density are tested, from 0.1 to 1.

As shown in Fig. 4.20, when the network density is increased, not only are more individuals infected overall, more individuals are infected in a shorter period of time, and the epidemic dies out quicker. This is due to the increased network density resulting in more individuals available to infect, causing the epidemic to spread more rapidly, and burnt out quicker. Of interest is the distributions of final epidemic sizes, as shown in Fig. 4.21. There seems to be two thresholds with regards to the network density. The first occurs between 0.1 and 0.3, where the epidemic goes from rarely infecting many individuals, to more often infecting the majority of the population. Secondly, between 0.7 and 0.9, the epidemic goes from infecting the majority of the population if it establishes itself, to almost every epidemic that establishes itself infecting every individual in a population.

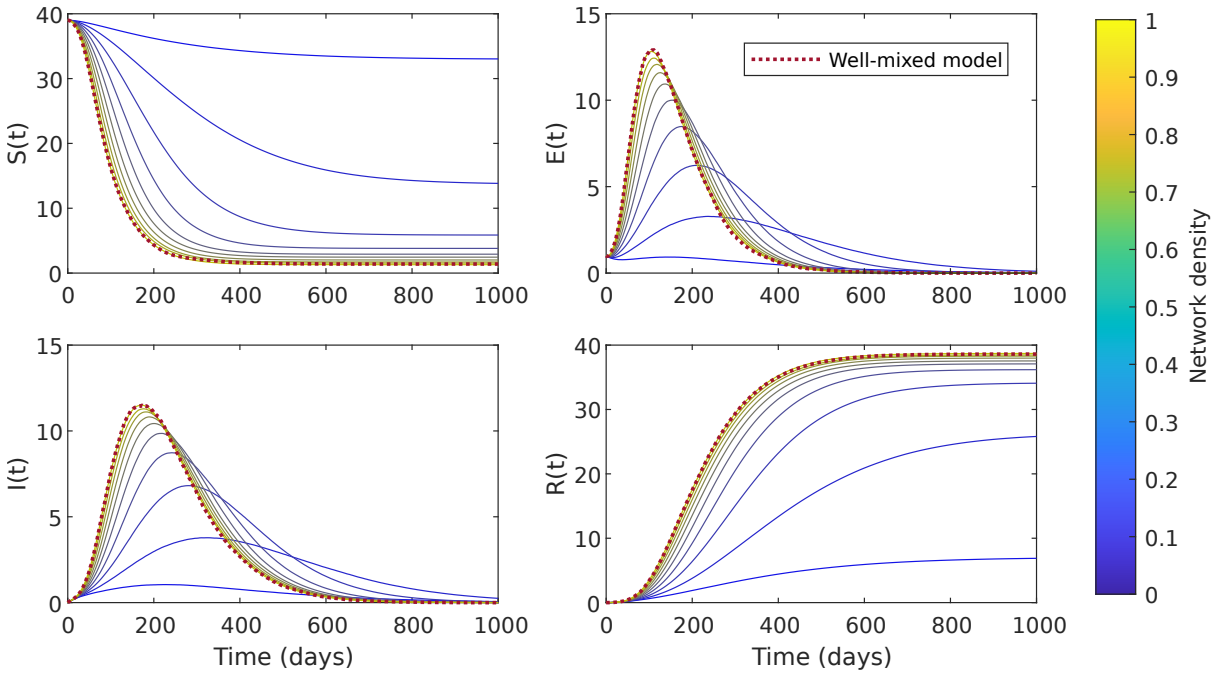



Figure 4.20: **Numbers of individuals in each compartment, for each 24-hour period, for ten different network densities.** As the network density is increased, more individuals are infected overall, more individuals are infected in a shorter period of time, and the epidemic dies out quicker. When the network density reaches 1, the results match those from the well-mixed model discussed in Section 4.3.

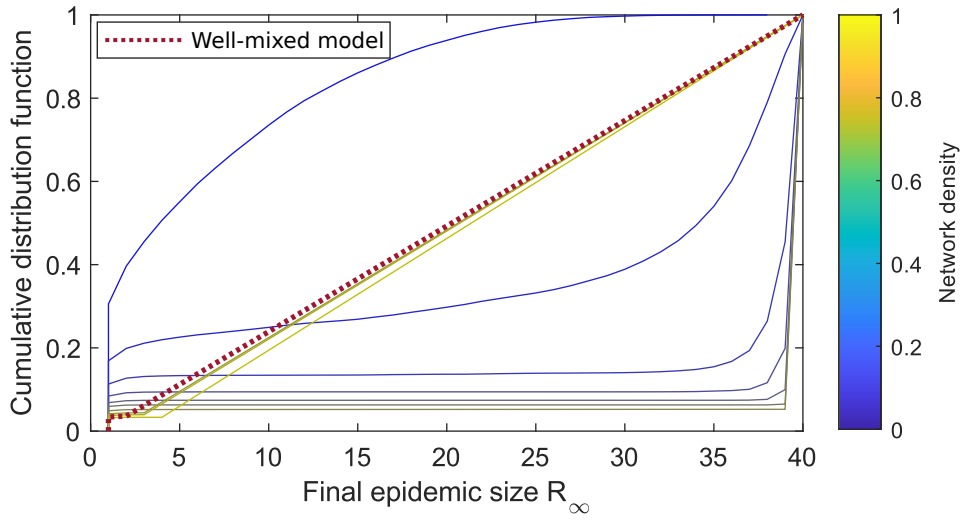



Figure 4.21: **Cumulative distribution functions of the final epidemic sizes for different network densities.** Two threshold situations appear. The first occurs between 0.1 and 0.3, where the epidemic goes from rarely infecting many individuals, to more often infecting the majority of the population. Secondly, between 0.7 and 0.9, the epidemic goes from infecting the majority of the population if it establishes itself, to almost every epidemic that establishes itself infecting every individual in a population. When the network density reaches 1, the results match those from the well-mixed model discussed in Section 4.3.

The distributions of final infection event times are shown in Fig. 4.22. Here, we can see that as the network density increases, the entire distribution at first (between 0.1 and 0.3) is shifted to the right, after which the peak is shifted to the left. As the network density increases, the number of routes of infection also increase, causing a higher number of infections, as shown in Fig. 4.21, but also for the epidemic to burn itself out quicker, as it runs out of individuals or infection routes.

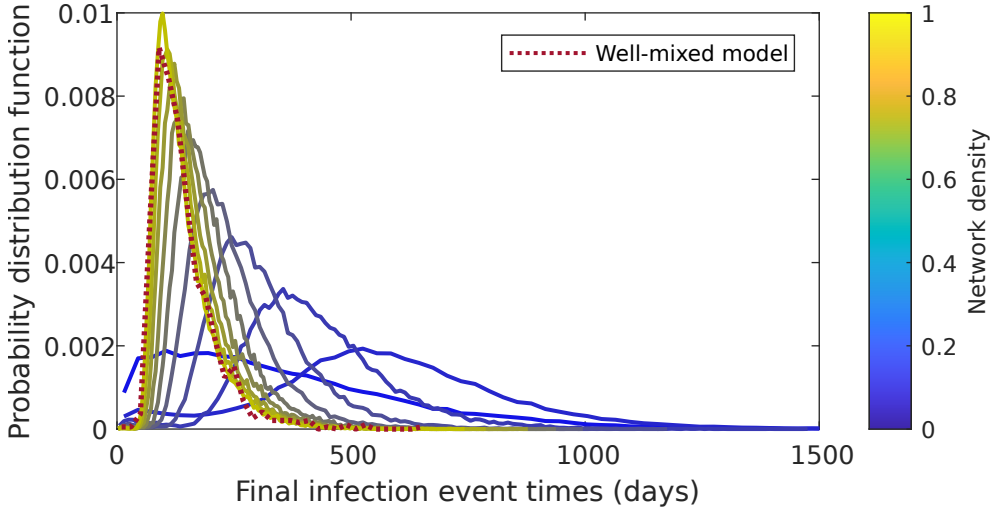



Figure 4.22: **Distributions of final infection event times for different network densities.** As the network density increases, the entire distribution at first (between 0.1 and 0.2) is shifted to the right, after which the peak is shifted to the left. The number of routes of infection increases along with the network density, meaning that not only are more individuals infected, but the epidemic burns itself out quicker, as it runs out of individuals or infection routes. When the netowkr density reaches 1, the results match those from the well-mixed model discussed in Section 4.3.

Plots of the distributions of final removal event times are given in Fig. 4.23. As the network density increases, the distribution changes in the same manner as with the final infection event times, with one added feature: the number of small event times decreases. This is most likely due to the increased network density allowing for more successful epidemics, which in turn last for a longer time.

When the network density is increased, the distribution of the number of infections from the seed case is shifted to the right, and the mean value, $R_0$, is increased, as shown in Fig. 4.24. An explanation could be that a higher network density, resulting in a higher mean degree, results in more individuals available, to the seed case, for infection.

When the network density approaches 1, each of the results that we examine match those from the well-mixed model discussed in Section 4.3. This is because the well-mixed model is equivalent to simulating epidemics on a complete network, that is one where all pairs of individuals are in contact.

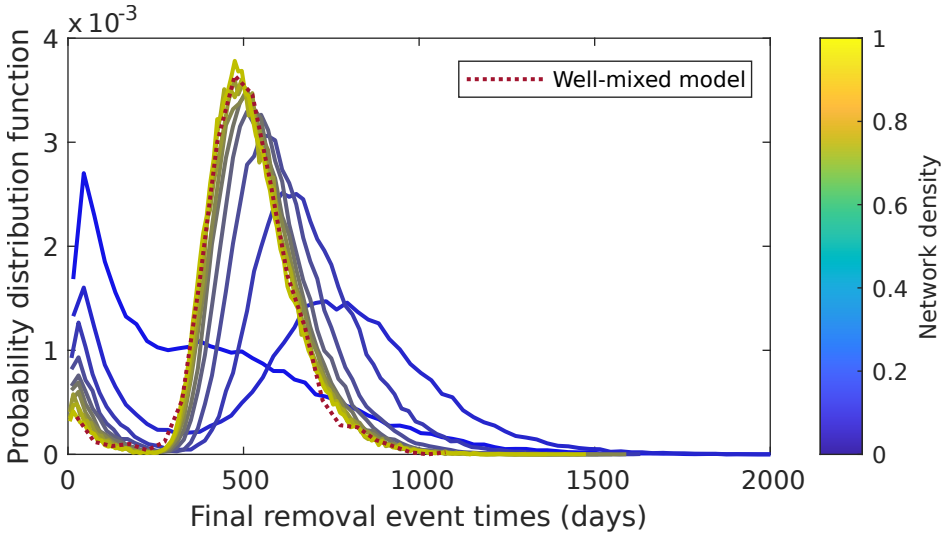



Figure 4.23: **Distributions of final removal event times for different network densities.** As the network density increases, the distribution changes in the same manner as with the final infection event times. However, an extra phenomenon appears. As the network density increases, the number of small final removal event times decreases. An explanation for this is the increased network density allowing for more successful epidemics, which in turn last for a longer time. When the network density reaches 1, the results match those from the well-mixed model discussed in Section 4.3.

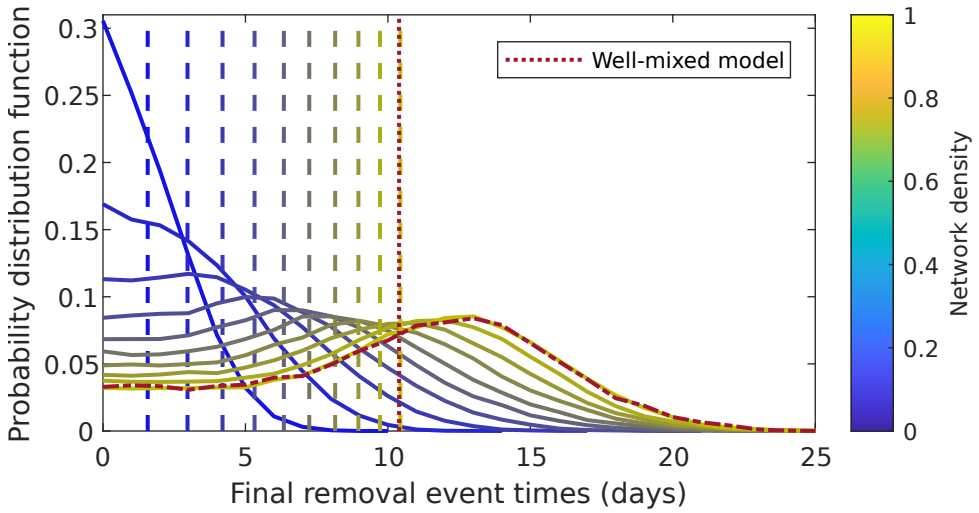



Figure 4.24: **The value of $R_0$ increases as the network density increases.** When the network density is increased, the distribution of the number of infections from the seed case is shifted to the right, and the mean value of these distributions, $R_0$, is increased. These values are shown as dashed lines. An explanation could be that a higher network density, resulting in a higher mean degree, results in more individuals available, to the seed case, for infection. Results from the well-mixed model are shown for comparison. As the network density reaches 1, the results from our stochastic transmission model match up with the well-mixed model.

### 4.7.3 Changing the choice of seed population

It is important, from both a theoretical and a real-world view point, to understand the effect upon a epidemic of the characteristics of the initially infected (either accidentally or deliberately) individuals. For example, if we desired to make sure that an epidemic spread as far through the population as possible, it would advantageous to know who which individuals would be best to infect in order to achieve this.

In this section, the effect of the process by which the seed population is determined is examined. For a fixed network density of 0.2, and parameters $\beta = 0.01$, $\gamma = \frac{1}{60}$, and $\delta = \frac{1}{75}$, we investigate three scenarios for the seed population: 'low-social' individuals, those whose degrees are less than two from the $25^{th}$ percentile, those with degrees less than two from the mean degree, and 'high-social' individuals, those whose degrees are less that two from the $75^{th}$ percentiles. These percentiles and mean are calculated from idealised random geometric graphs on 40 individuals, and plots of these values for various densities are shown in Fig. 4.25.

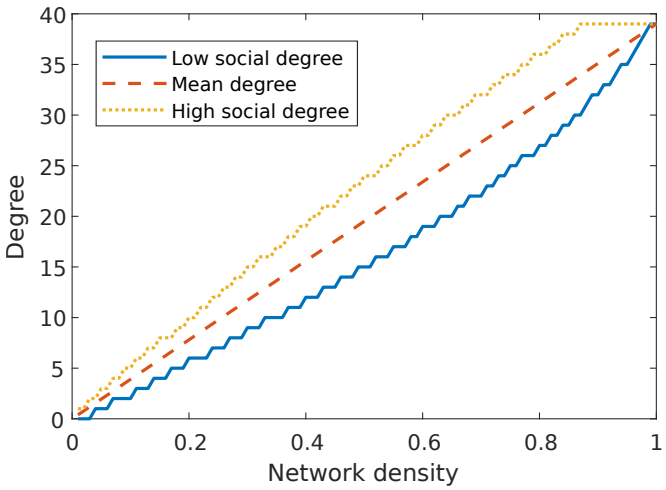



Figure 4.25: Plots of the $25^{th}$, $50^{th}$ (mean), and $75^{th}$ percentiles of the degrees of idealised random geometric graphs on 40 individuals.

Altering the seed population has an effect on the epidemic simulations that we perform, though the effect is lesser than when the parameters $\beta$, $\gamma$, and $\delta$ are varied. Seeding the epidemic in more social individuals results in a higher final epidemic size (as shown in Fig. 4.27), higher numbers of exposed and infected individuals and the epidemic ending marginally quicker (as shown in Fig. 4.26). Further, the time of both the final infection and final removal events occurs earlier if a more social individual were chosen (as shown in Figs. 4.28 and 4.29). Finally, the value of $R_0$ is higher if a more social individual were chosen, due tpo the higher number of individuals that can be infected. This is shown in Fig. 4.30.

Changing the seed population does allow the results from the stochastic transmission model to converge to those from the well-mixed model. As shown in Sections 4.7.1 and 4.7.2, unless the values of the epidemic parameters or network density are chosen accordingly, the results from the stochastic transmission and well-mixed models will not converge.

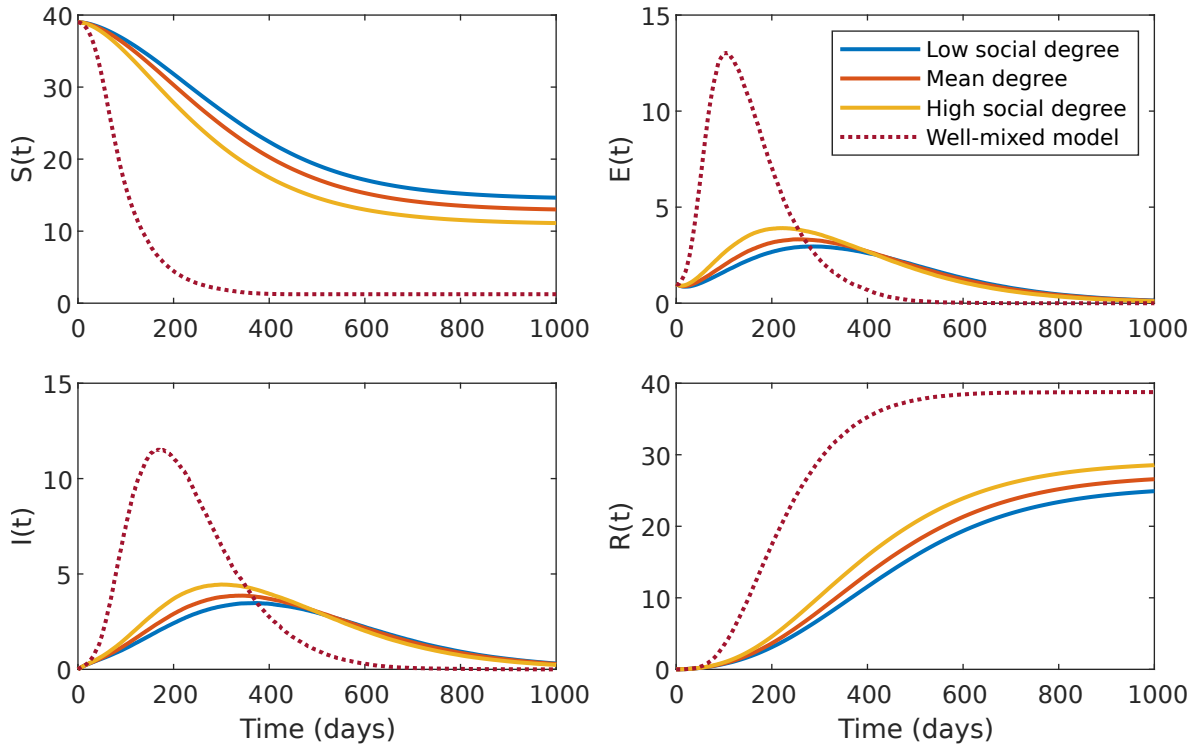



Figure 4.26: **Numbers of individuals in each compartment, for each 24-hour period, for three different seed population choices.** When the seed population is chosen from individuals with higher degrees, the number of removed individuals increases, as does the number of exposed and infected individuals, and the epidemic ends quicker.

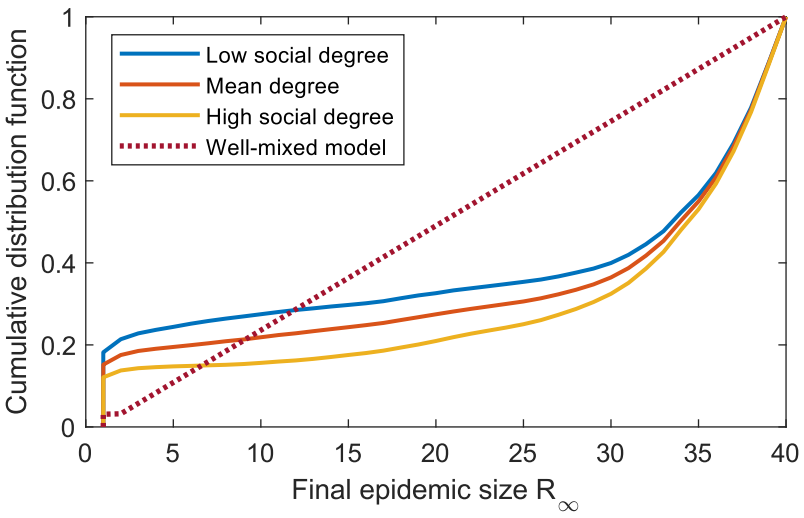



Figure 4.27: **Cumulative distribution functions of the final epidemic sizes for three different seed population choices.** As the degree of the seed case increases, so does the final epidemic size $R_\infty$. This is a result of more individuals being infected in the first stages of the epidemic, allowing it to spread further.

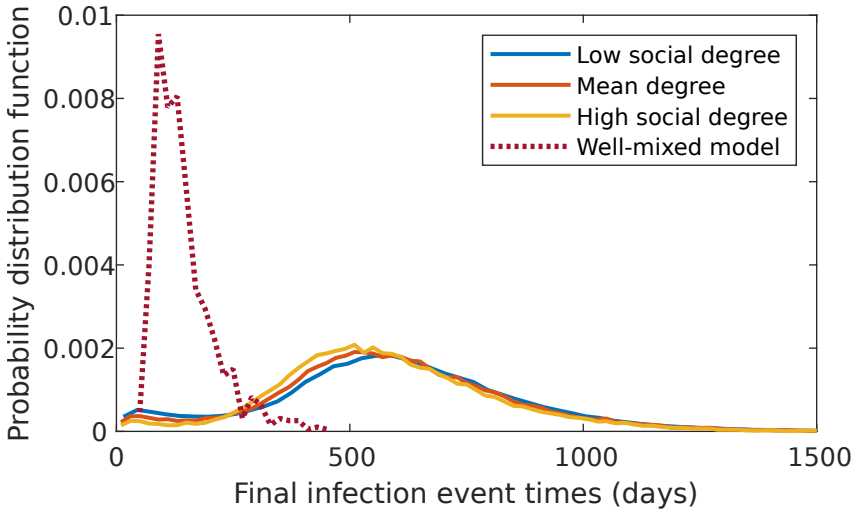



Figure 4.28: **Distributions of final infection event times for three different seed population choices.** As the degree of the seed population increases, the final infection events occur earlier.

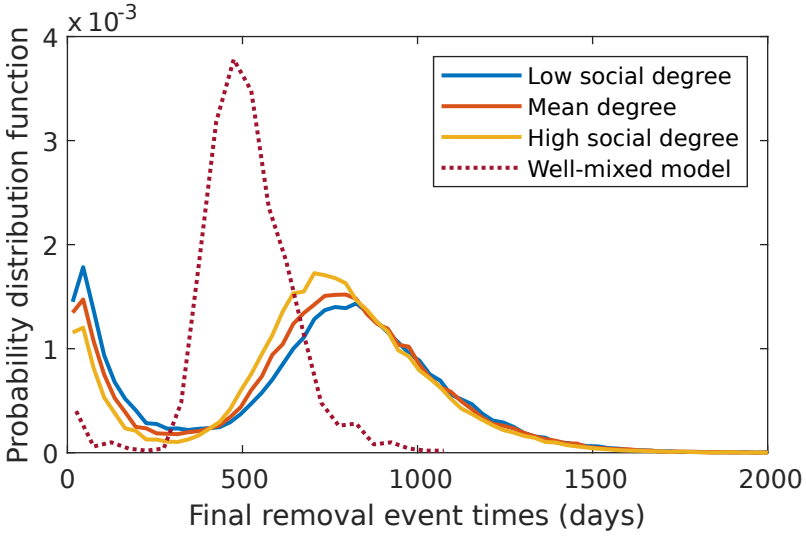



Figure 4.29: **Distributions of final removal event times for three different seed population choices.** As the degree of the seed population increases, the final removal events occur earlier.

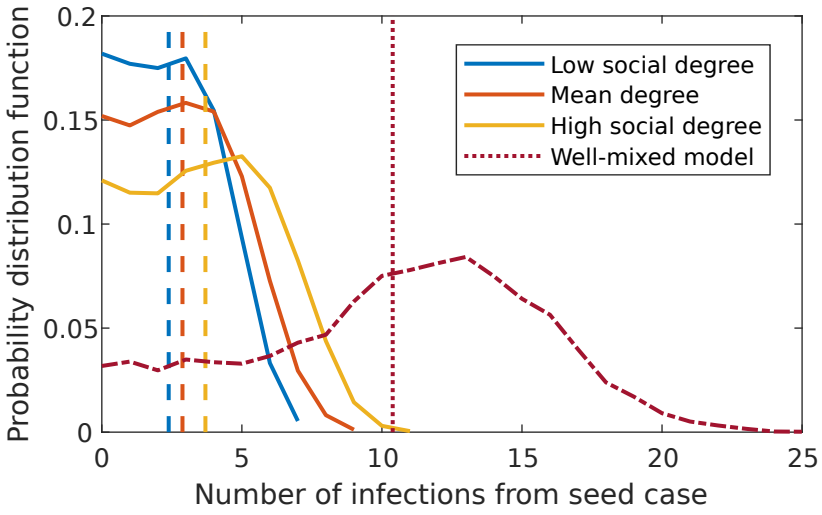



Figure 4.30: **The value of $R_0$ choice of seed case has a small effect on the value of $R_0$.** Histograms of the number of infections directly transmitted from the seed case are calculated from simulations. The values of $R_0$ are shown as dashed lines. The results from the well-mixed model are shown for comparison.

### 4.7.4 Changing the distribution of edge weights

In the following chapter, we present network models for the spread of bovine tuberculosis through possum populations. The networks upon which epidemics are simulated are weighted, so it is necessary to understand how changing the distribution of edge weights affected epidemic spread.

To achieve this, we fix the parameters and network density as before. Then we investigate three scenarios: positively skewed edge weights, unweighted edges (binary networks), and negatively skewed edge weights. Each scenario is designed to have a mean edge weight of 1. In the positively (negatively) skewed scenario, the weight of an edge between nodes $i$ and $j$ is an increasing (decreasing) linear function of the distance between them (see Fig. 4.31). Histograms of positively and negatively skewed edge weights produced from 2000 populations, each of network density 0.2 and size 40, are shown in Fig. 4.32. The means of these distributions are represented by lines, and show that the means are approximately one.

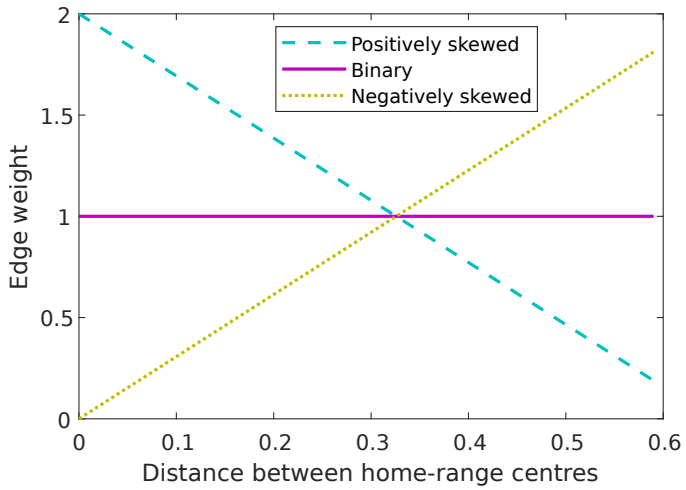



Figure 4.31: Plots of functions of distance between home-range centres that give edge weights.

The numbers of individuals in each compartment, for each edge weighting scenario are shown in Fig. 4.33. As the edge weights become more negatively skewed, the number of susceptible individuals decreases, and there are more infections events in a given period of time. This is corroborated by the distributions of final epidemic sizes shown in Fig. 4.34. However, as shown in Figs. 4.35 and 4.36, the edge weightings have a comparatively minor effect on the distributions of final infection and final removal event times, which could be due to the edge weights being too similar. As illustrated in Fig. 4.37, the edge weights have little effect on the value of $R_0$. This could be due to the edge weights being too similar.

In each of the three edge weight distribution scenarios, the results from the stochastic transmission model do

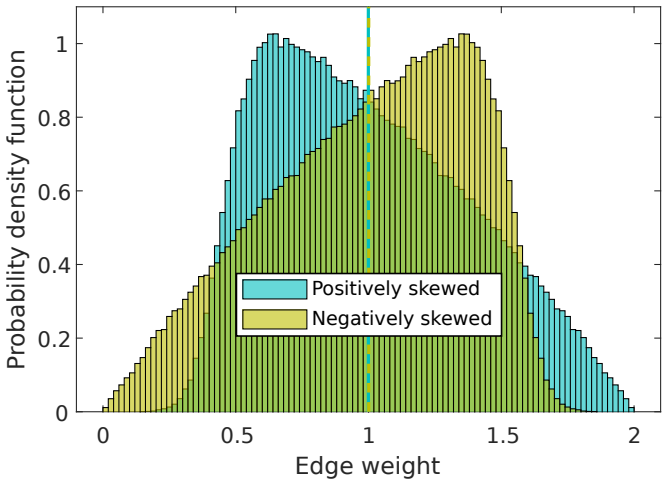



Figure 4.32: Distributions of positively and negatively skewed edge weights produced from 2000 populations, each of network density 0.2 and size 40. The mean values of the distributions are shown as lines.

not match up with those from the well-mixed model. This is most likely due to the edge weights being very similar and tightly distributed about one. As discussed in Section 4.7.3, unless the parameters are chosen accordingly, the results will not converge. Furthermore, as the edge weightings are very similar to that of the networks used in these aforementioned sections, the results will once again not converge.

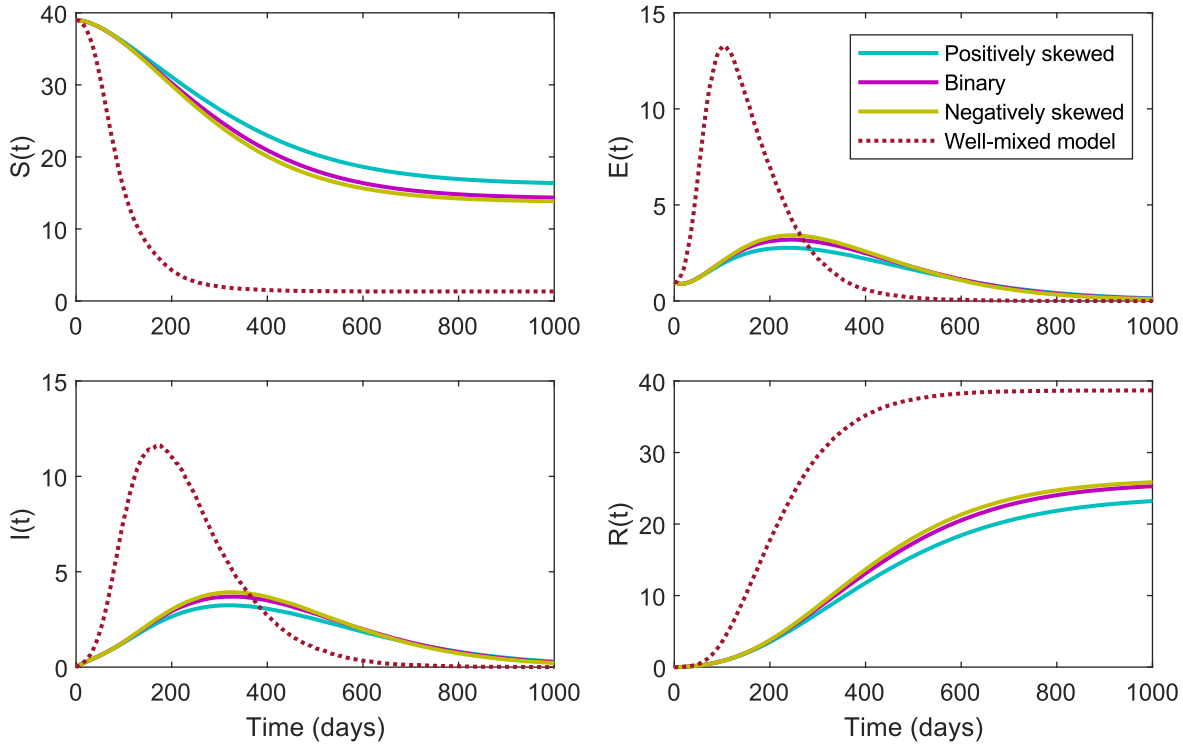



Figure 4.33: **Numbers of individuals in each compartment, for each 24-hour period, for different edge weightings.** As the edge weights become more negatively skewed, the number of susceptible individuals decreases, and there are more infections events in a given period of time.

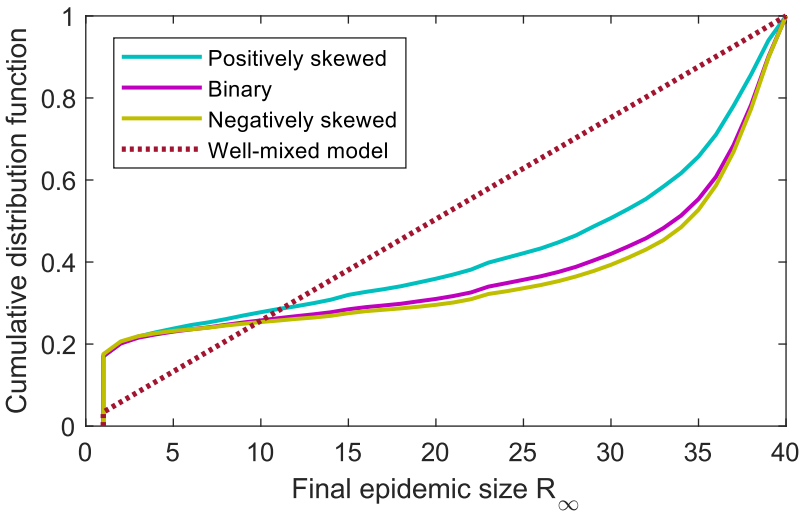



Figure 4.34: **Cumulative distribution functions of the final epidemic sizes.** The numbers of individuals in each compartment, for each edge weighting scenario are shown in Fig. 4.33. As the edge weights become more negatively skewed, the number of susceptible individuals decreases, and there are more infections events in a given period of time. This is corroborated by the distributions of final epidemic sizes shown in Fig. 4.34.

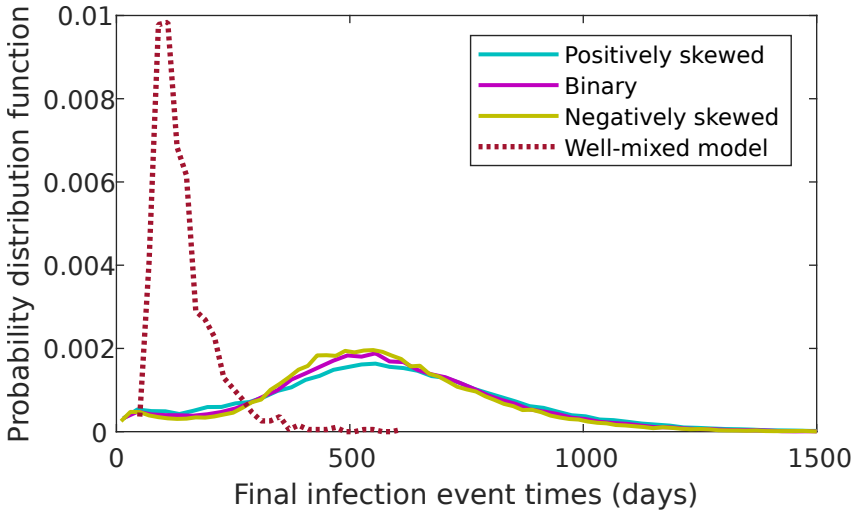



Figure 4.35: **Distributions of final infection event times for different edge weighting scenarios.** The distributions are similar, regardless of the edge weighting. This is possibly due to edge weights being very similar.

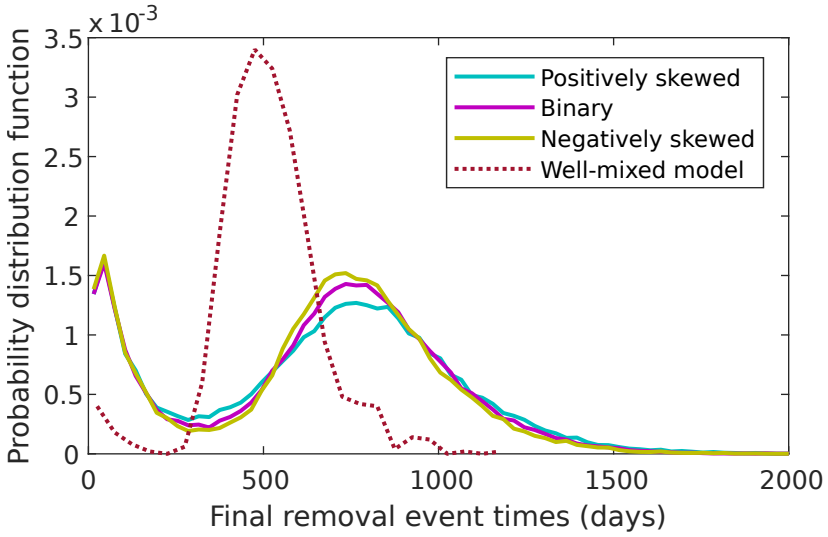



Figure 4.36: **Distributions of final removal event times for different edge weighting scenarios.** The distributions are similar, regardless of the edge weighting. This is possibly due to edge weights being very similar.

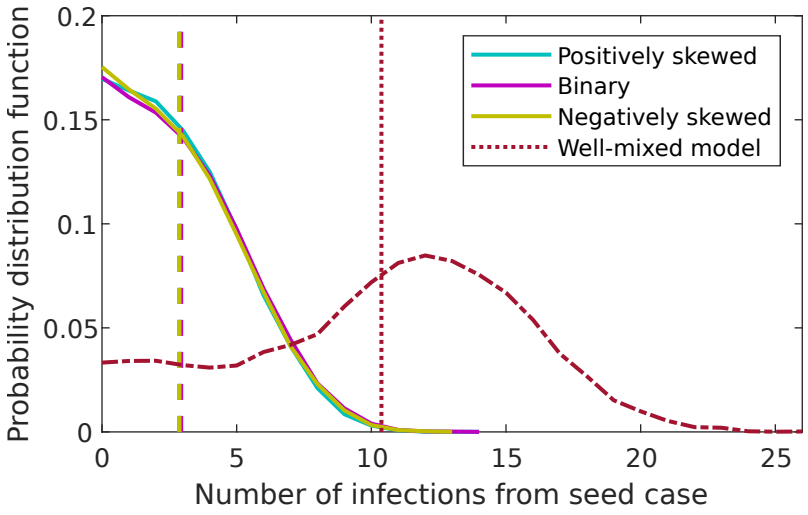



Figure 4.37: **The edge weighting has little effect of the value of $R_0$.** Distributions of the number of infections caused by the seed case are calculated from simulations. The distributions are similar, indicating that the edge weight function shown in Fig. 4.31 has little effect of the values. Accordingly, the mean of the distributions, $R_0$, is also similar. These values are shown as dashed lines. Results from the well-mixed model are shown for comparison.

## 4.8 Discussion

In this chapter, we presented a stochastic transmission model for simulating the spread of an epidemic through a population represented by a network structure.

We started by examining the classical deterministic well-mixed compartmental model, where individuals are assigned to compartments according to their epidemic status and can move between compartments according to given transition rates. The well-mixed model also assumes that all individuals are in contact and make contact at the same rate. We highlighted important results from this model, including those concerning the final epidemic size $R_\infty$ and basic reproductive number $R_0$.

Following on from the well-mixed model, we discussed our stochastic transmission model. Performed on a weighted contact network, we use an SEIR-model to simulate infection events between susceptible and infectious individuals, in addition to events representing exposed individuals becoming infectious and infectious individuals being removed from the population. From a given simulation, we can produce a variety of results about the spread of a disease through a network. These results include the numbers of individuals in each compartment (Susceptible, Exposed, Infectious, and Removed) at the end of each 24-hour period, the final epidemic size $R_\infty$, and the basic reproductive number $R_0$.

Subsequently, we compared our stochastic transmission model against both the deterministic and a stochastic version of the well-mixed model. We found that simulating the stochastic transmission model on a complete network allows the results to match up with the stochastic well-mixed model, but neither matched with the deterministic well-mixed model. This is most likely due to stochasticity allowing for scenarios including a non-zero probability of the epidemic dying out. We also examined analytical results for the basic reproductive number $R_0$, comparing them against values produced from simulations. We found that analytical results were consistently greater than the simulated values, indicating that our process of calculating $R_0$ from simulations is potentially erroneous.

In the last part of this chapter, we used simulations on toy networks to investigate the effect of altering: parameters including the transmission coefficient $\beta$, the rate of becoming infectious $\gamma$, and the removal rate $\delta$; network characteristics including network density and edge weights; and the initial condition of the seed

population. We found that overall, changing either the epidemic parameters or the network density had the largest effect on the evolution of an epidemic. When increasing $\beta$ or the network density, or decreasing $\delta$, not only are more individuals infected, but specifically more individuals are infected by the seed case. Further, when increasing $\beta$, $\gamma$, or the network density, the epidemic dies out sooner. Effects can also be seen when altering the edge weights, or seeding the infection in individuals with a higher or lower than mean degree, but the effects are much less stark.

When considering the well-mixed model, in comparison to our stochastic transmission model, we found that when increasing the transmission coefficient $\beta$ to a value above 0.1, or increasing the network density to close to 1, the two methods converge. In other parameter, network topology and edge weight situations, this does not hold.

The model we have presented in this chapter will be used in the following chapter to model the spread of a bovine tuberculosis epidemic through a common brushtail possum population. We simulate epidemics on the empirical contact networks from Chapter 2, in addition to networks produced from Chapter 3, with a view to find appropriate parameter values that will allow results from our transmission model to match with those from the real-world data discussed in Section 2.3.

## 4.9 Glossary of notation used

| Notation | Meaning |
|---|---|
| $N$ | Population size. |
| $S$, $E$, $I$, $R$ | States of individuals in the population. |
| $S(t)$, $E(t)$, $I(t)$, $R(t)$ | Number of individuals in each state at time $t$. $S(t) + E(t) + I(t) + R(t) = N$. |
| $i$, $j$ | Individuals in our population. |
| $\mathbf{x}(t)$ | State of the population at time $t$. The initial state of the population is $\mathbf{x}(0)$. |
| $x_i(t)$ | State of individual $i$ at time $t$. Here, $x_i(t) \in \{S, E, I, R\}$. |
| $\beta$ | Transmission coefficient. |
| $\gamma$ | Rate at which individuals who are exposed become infectious. |
| $\delta$ | Rate at which infectious individuals are removed. |
| $q_k(t)$ | Transition rate for transition $k$ at time $t$. |
| $q(t)$ | Total transition rate for the system at time $t$. Here $q(t) = \sum_{k=1}^{N} q_k(t)$. |
| $P_{ij}$ | Probability of a contact occurring between individuals $i$ and $j$. |
| $\kappa_i(t)$ | Rate at which individual $i$ is exposed to the disease at time $t$. |
| $R_0$ | Basic reproduction number. |
| $R_\infty$ | The final size of the epidemic. |
| $\mathcal{D}$ | Diffusion coefficient. |

Table 4.3: Table of notation for Chapter 4

# Chapter 5

# Fitting a stochastic transmission model for the spread of bovine tuberculosis through a possum population to real-world data

## 5.1 Background

In Chapter 2, we examined two data sets, one of contacts between possums in four distinct populations, and one of the introduction and later detection of bovine tuberculosis in these same populations. As discussed in Section 2.3, secondary infections of individuals who were never collared were detected. Because of the absence of contact data regarding these uncollared secondary infections, it is necessary for us to produce a model describing the spread of a bovine tuberculosis epidemic through a possum population composed of both collared and uncollared individuals. A model of possum movement and contact was presented in Chapter 3, which allowed us to produce weighted contact networks representing possum populations. In Chapter 4, we presented a stochastic transmission model for the spread of an epidemic through a population represented by a weighted network.

This chapter brings together much of the work conducted in these preceding chapters to present an epidemic

model for the spread of a bovine tuberculosis epidemic through a possum population. This has been modelled previously, for example by Barlow in both 1991 and 2000, as well as Roberts in 1992 and 1996. Whilst these models are well-established, they differ from the approach that we take in three keys areas. The first is the contact structure of the population. In Barlow's and Robert's models, an assumption of population mixing is made, where each possums has a chance of coming into contact with any other possum. However, that is not necessarily realistic, as in our data we have seen that many possum never meet. We instead follow a similar approach to that of Ramsey and Efford (2010), where possums are represented by points in space and the infection transmission rate is a function of the distances between these points.

Second is the deterministic nature of the models. Barlow (1991a), Roberts (1992), Roberts (1996) and Barlow (2000) all present deterministic models, whereas ours is instead fundamentally stochastic. The final key area is population dynamics. In most preceding models, there is an assumption of birth, non-disease related death, migration, and pseudovertical transmission. However, we do not incorporate these because on the timescale we model they have less bearing on the results.

The method of using contact networks for understanding the spread of an epidemic through a real-world population has been applied to many different diseases and contexts, including the spread of tuberculosis. Drewe et al. (2011) for example, simulates an SEIR-model of TB spread in meerkats on empirical weighted contact networks produced from contact data. Corner et al. (2003) and Porphyre et al. (2008) both investigated the effect of contact structure on TB spread in possums, however, they did not utilise compartmental models in their papers.

After describing the model, we utilise the secondary infection data discussed in Section 2.3, to estimate the transmission coefficient $\beta$. Using the estimated values of $\beta$, we perform simulations on model networks and discuss the resulting epidemics, finding that our stochastic network model is able to replicate certain aspects of the disease spread.

## 5.2 Stochastic network model of bovine tuberculosis spread in common brush-tail possums

In this section, we present a stochastic network model for the simulation of a bovine tuberculosis epidemic through a population of model individuals. We use the compartmental SEIR-model described in Section 4.4. Recall from Chapter 4, that the following assumptions are made about the model population:

1. The rates of birth and non-disease induced mortality are zero, resulting in the population size, $N$, remaining constant.

2. The population is not well-mixed, so it is represented by a weighted contact network $G$, where the edge between individuals $i$ and $j$, if it exists, is weighted by $w_{ij}$, the contact rate.

3. Infection is transmitted from infectious individuals to susceptible individuals by contact. The rate at a contact results in transmission is given by the transmission coefficient $\beta$.

4. Exposed individuals become infectious, and infectious are removed from the population, at constant rates $\gamma$ and $\delta$ respectively. Removed individuals do not make any further contact with other individuals, and therefore cannot transmit the infection.

Simulations of an epidemic spreading through a weighted contact network, representing a possum population, are performed as per Section 4.4. There are three aspects to our possum epidemic model that need to be considered in further detail: the creation of the weighted contact network, the choice of seed population, and determination of the epidemic parameters used. We examine each separately.

### 5.2.1 Producing weighted contact networks

We create a population of individuals, representing possums, using the possum home-range centre distribution model from Section 3.4.1. Here, a population of $N$ individuals, are placed in a circular region of radius $R_I$ according to a spatial Poisson point process with density parameter $\lambda > 0$. A subset of these individuals, of size $C$, are then collared according to the Uniform Collaring Model described in Section 3.4. The parameters used to produce the populations are given in Table 5.1. The Uniform Collaring Model was chosen because it performed best when estimating the distribution of nightly contact probability edge weights, as discussed in Section 3.7.

A binary network is produced from this model population, according to the method described in Section 3.6, where each individual is associated to a node, located at the individual's home-range centre. The pairwise distances between nodes are calculated, and values of the nightly contact probability $P_{ij}$ (described in Section 3.3.1) are calculated using the mean optimal diffusion coefficient estimated in Section 3.5. The diffusion coefficients values that are used are given in Table 5.1. Edges exist between the nodes representing individuals $i$ and $j$ if the value of $P_{ij}$ is greater than the threshold value $1/190$, chosen to reflect the minimum number of nights that a pair of possums in the real-world data discussed in Chapter 2 could have met. Fig. 5.1 shows an example model population, produced using the parameters for site A. The light grey nodes represent individuals who are trapped but not collared, and the dark grey nodes represent collared individuals.

We weight the edges of this network by the contact rate $w_{ij}$. The value of $w_{ij}$ is calculated as the expected number of contacts per 24 hour period. We assume that possums do not contact each other during the day, and therefore the number of diurnal contacts is zero. This assumption is unrealistic for two important reasons. The first is that there exists diurnal data in the real-world data set discussed in Section 2.2, indicating that daytime contact between possums took place, either through meeting one another, or by den-sharing. The second reason is that, whilst contact may play a part in transmission of bovine tuberculosis, den-sharing will also have an impact, and may even have a greater impact due to the close proximity and length of time spent asleep increasing the chances of transmission. Despite those reasons, this assumption was made because our model of possum movement and contact does not model the diurnal behaviours of possums. Therefore we have no justification for any of the possible modelling mechanisms that would include daytime contacts.

| Site | N° possums trapped $N$ | N° possums collared $C$ | Estimated density $\lambda$ ($m^{-2}$) | Inhabiting area radius $R_I$ (m) | Mean optimal diffusion coefficient $\mu_{\mathcal{D}}$ ($m^2s^{-1}$) | Estimated transmission coefficient $\beta$ |
|---|---|---|---|---|---|---|
| A | 142 | 35 | $4.87\times10^{-4}$ | 308.91 | 0.03881 | - |
| B | 174 | 37 | $6.92\times10^{-4}$ | 286.14 | 0.01162 | 0.02307 |
| C | 144 | 40 | $4.08\times10^{-4}$ | 335.18 | 0.07004 | 0.00826 |
| D | 140 | 39 | $4.20\times10^{-4}$ | 325.74 | 0.02986 | 0.02018 |

Table 5.1: **Summary data of the possum populations**. Values, pertaining to the real-world possum populations discussed in Chapter 2. The values of $N$, $C$, $\lambda$ are taken directly from Richardson et al. (2017), whereas $R_I$ is calculated as $R_I = \sqrt{N/\pi\lambda}$). The mean optimal diffusion coefficient is drawn from Table 3.2. On sites A and B, there are four less individuals than when these sites were examined in Chapter 3. This is due to the four possums who were experimentally challenged but died soon afterwards, see Section 2.1. The values of the transmission coefficient are estimated by comparing simulated results to the real-world secondary infection numbers.

Regarding nightly contacts, from Fig. 2.9 we have that the modal number of contacts per night is one. Multiplying the modal number of contacts by the expected number of nights during which a contact will occur, given by $P_{ij}$, we have that the expected number of nightly contacts in equal to $P_{ij}$. As we have assumed no diurnal contacts, the expected number of contacts per 24 hour period is given by $P_{ij}/2$, and so $w_{ij} = P_{ij}/2$.

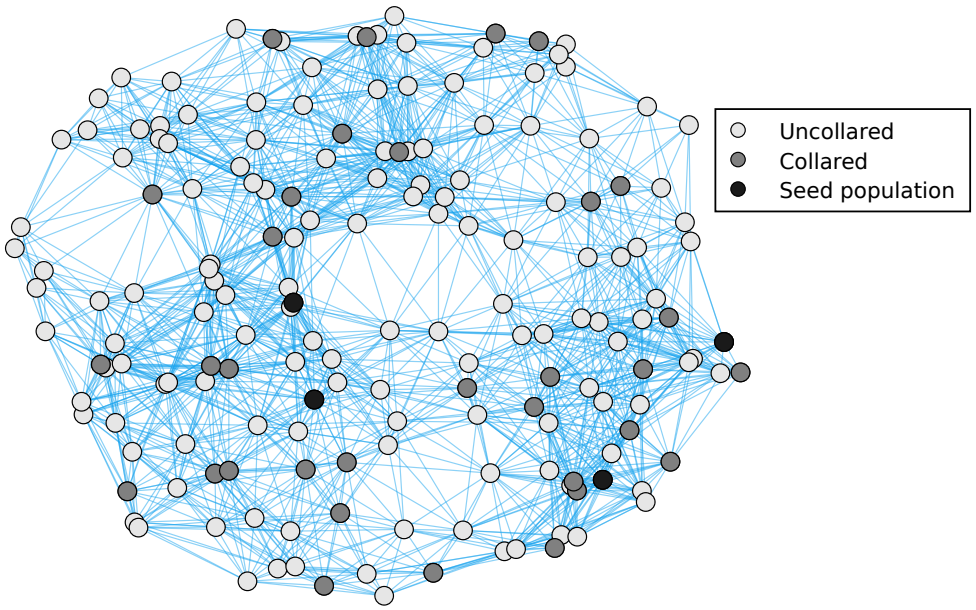



Figure 5.1: **Example model population, produced using the parameters, given in Table 5.1, for site A.** The light grey nodes represent individuals who are trapped but not collared, and the dark grey nodes represent collared individuals.

### 5.2.2 Choosing the seed population

As discussed in Section 4.4, the seed population is composed of a small proportion of collared individuals who are initially exposed to the disease at the beginning of the epidemic simulation. In the real-world population whose data is discussed in Section 2.3, four individuals on each of the four study sites were experimentally infected with bovine tuberculosis.

When performing epidemic simulations, we aim to match up the real-world seed populations with individuals in our model population who are similarly connected in their respective network. To do this we perform the following steps:

1. Calculate the degrees of each individual's node in both the empirical (or real-world) and model networks.
2. Rank the nodes in the empirical network, and in the model network, from highest degree to lowest degree. This gives two sorted lists of vectors.
3. Choose nodes from the model network by selecting those nodes with the same ranking as the empirical nodes who were part of the real-world seed population.

An example of this process is given in Fig. 5.2, where nodes for site B and a model population of collared individuals are matched up by degree. The black nodes are the seed population individuals.

The rationale for this process is that, whilst it could be more appropriate to match up an empirical node of degree three with a spatial node of the same degree, there is no guarantee that there will always be. The nodes of the same degree are permuted to remove any bias brought about by a sorting algorithm. A disadvantage to this process, as illustrated by Fig. 5.2, is that the empirical nodes are not always matched with nodes of similar degree.

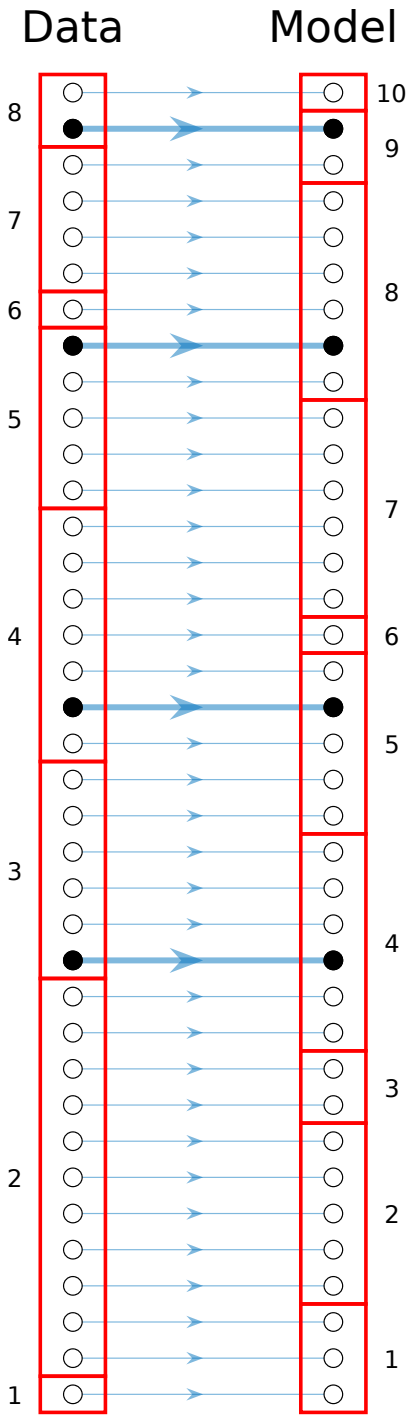



Figure 5.2: **An illustration of the node matching process for site B, as described in Section 5.2.2.** The nodes are ranked by degree, then those in the empirical data populations that were experimentally infected are matched to corresponding individuals in the model population.

### 5.2.3 Determining the appropriate values of the epidemic parameters $\gamma$ and $\delta$

When transitioning from a well-mixed SEIR-model to a network-based model, the rates of becoming infectious $\gamma$ and of removal $\delta$ do not need to be adjusted. This is because the movements of individuals between the exposed and infectious, and the infectious and removed, compartments is not affected by the underlying network structure. In common brushtail possums, clinical signs of TB infection begin to show 2-3 months post infection [Roberts, 1992, Corner et al., 2002], and possums typically die within 4-5 months of clinical signs appearing [Ramsey and Cowan, 2003]. Whilst possums may be infectious before clinical signs appear, we have chosen to use values of $\gamma = 0.0133$ and $\delta = 0.0074$, which are equivalent to an average latent period of 75 days, and average infectious period of 135 days. However, the same does not hold true for the transmission coefficient $\beta$.

### 5.2.4 Estimating the transmission coefficient $\beta$

The well-mixed compartmental model assumes that all individuals are coming into contact at the same rate. We do not make this assumption, so the value of the transmission coefficient $\beta$ will be sensitive to the specific contact structure of the population. The process by which $\beta$ is calculated for network populations is less clear, as the transmission of disease between individuals is directly affected by the population's contact structure.

Barlow in his 2000 paper stated that "Disease transmission coefficients are notoriously hard to measure in the field", a conclusion that is even more applicable to our situation, with the calculation of $\beta$ being very sensitive to the specific contact structure of the population. Porphyre et al. (2008) calculated a lower estimate for the transmission coefficient, finding that above a value of 0.08 an epidemic will establish itself. However, their methodology is not totally applicable to us. Whilst they formed contact networks, the networks were unweighted. Additionally, the definition of a contact is different to ours, inferring contact based on where individuals were trapped.

We estimate values of $\beta$ by performing epidemic simulations on model networks, and comparing the epidemic evolution results to the real-world data discussed in Section 2.3. This process is carried out for a range of values of $\beta$, and the value that produces results closest to the real-world data is then chosen.

From our data (see Chapter 2), we have for each site, except site A, a singular secondary infection. Each

secondary infection occurred in an uncollared individual, whose infection could therefore not be explained. We produce 2000 populations, and perform 50 epidemic simulations on each population, with a different seed population in each simulation. Aggregating the results, we compare the number of secondary infections of uncollared individuals with the real-world data. To compare appropriately, we determined the number of secondary infections after the same amount of time had elapsed as between the initial experiment infections, and the detection of the secondary infections.

For values of the transmission coefficient $\beta$ between 0 and 0.05, we determined the error or difference between the simulated results and the real-world data. A plot of this error is given in Fig. 5.3. From this plot, we interpolate to find the value of $\beta$ on each site that gives an error of 0. These interpolated values are given in Table 5.1. The estimated values of the transmission coefficient are of the same order of magnitude as the value given by Porphyre et al. (2008), which indicates that our estimated values are sensible.

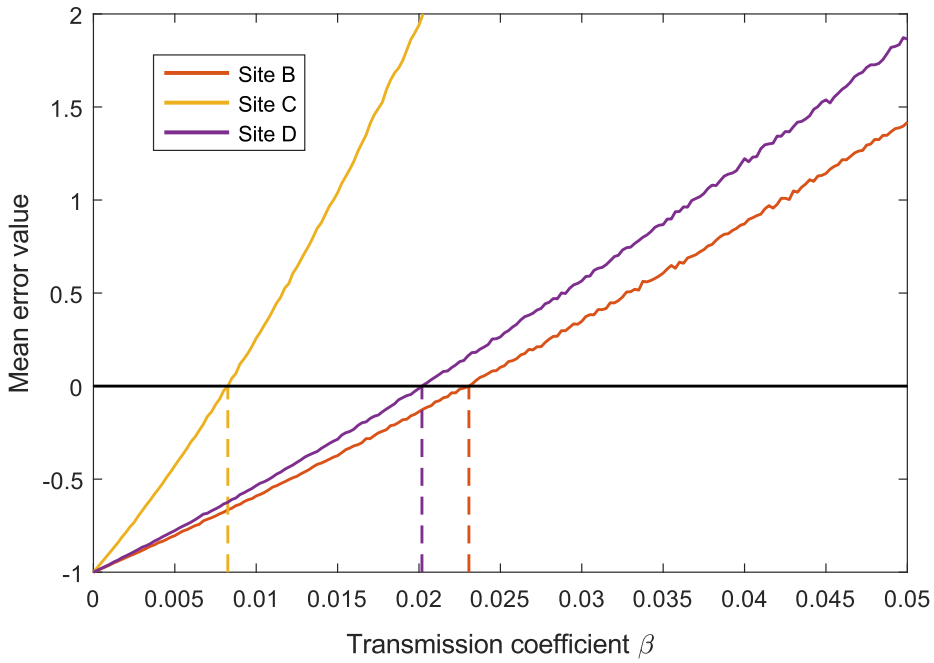



Figure 5.3: **Values of the difference between the simulated results and the real-world secondary infection numbers, for a range of the transmission coefficient $\beta$.**

### Summary

We have now presents our stochastic transmission model for the simulation of a bovine tuberculosis epidemic through a weighted network. We have detailed the process by which model populations are created, and how weighted networks are then produced. We have estimated values of the transmission coefficient $\beta$ by performing simulations and determining the value that produces simulated results closest to the real-world

data. By comparing with Porphyre et al. (2008), we can say that the values of $\beta$ that we estimate are sensible. These values will now used to simulate epidemics on model networks, allowing us to determine how effective our stochastic transmission model is at replicating the epidemic dynamics.

## 5.3 Results

We perform simulations on model weighted networks, using both the estimated transmission coefficient $\beta$, and a fixed value of $\beta = 0.0231$. This fixed value of $\beta$ was chosen because it is greater than each of the estimated values given in Table 5.1, resulting in the overstimation of the number of secondary infections of uncollared individuals. This is more realistic than choosing a value that would underestimate the number on a particular site, as in reality it is likely that secondary infection did occur, but were not detected.

For each site, and for both the estimated and fixed values of $\beta$, we produce 2000 populations, and simulate 50 epidemics on each. Each epidemic simulation has a different seed population. Site A does not have an estimated values of $\beta$, so we do note perform those corresponding simulations. Fig. 5.4 shows the average number of individuals in each compartment for each day.

Fig. 5.5 shows the number of non-collared secondary infections at the end of each days. When comparing to the secondary infections data (shown in the figure as a red dot), we can see that using the estimated values of $\beta$ allows for this number to be well-estimated. This is to be expected. When the larger, fixed value of $\beta$ is used, the number is overestimated, but not considerably.

Whilst $R_0$ has been calculated for TB in New Zealand-based possums a number of times, the respective models have assumptions that do not match with our model. Barlow (1991a), for example, does use an SEIR-model, but when calculating $R_0$ for the three different disease scenarios they include terms representing pseudo-vertical transmission between mothers and young, as well non-disease-induced mortality. No aspect of the explicit contact structure is included in the calculation of $R_0$. Roberts (1996) and Barlow (2000) use a different formula to calculate $R_0$, however theirs still lacks information about the explicit population contact structure, and includes population dynamic terms such as birth and non-disease-induced mortality. Overall, a range of values of $R_0$, from 1.13 to 2.13 have been calculated.

Fig. 5.6 shows the distribution of the number of secondary infections directly transmitted by the members of the seed population. As there are four individuals in each seed population, we divide by four in order to compare with the literature. The average number of secondary infections, that is the estimated $R_0$ value, is therefore approximately within the range 1.13 to 2.13 on each site, regardless of which values of $\beta$ is used.

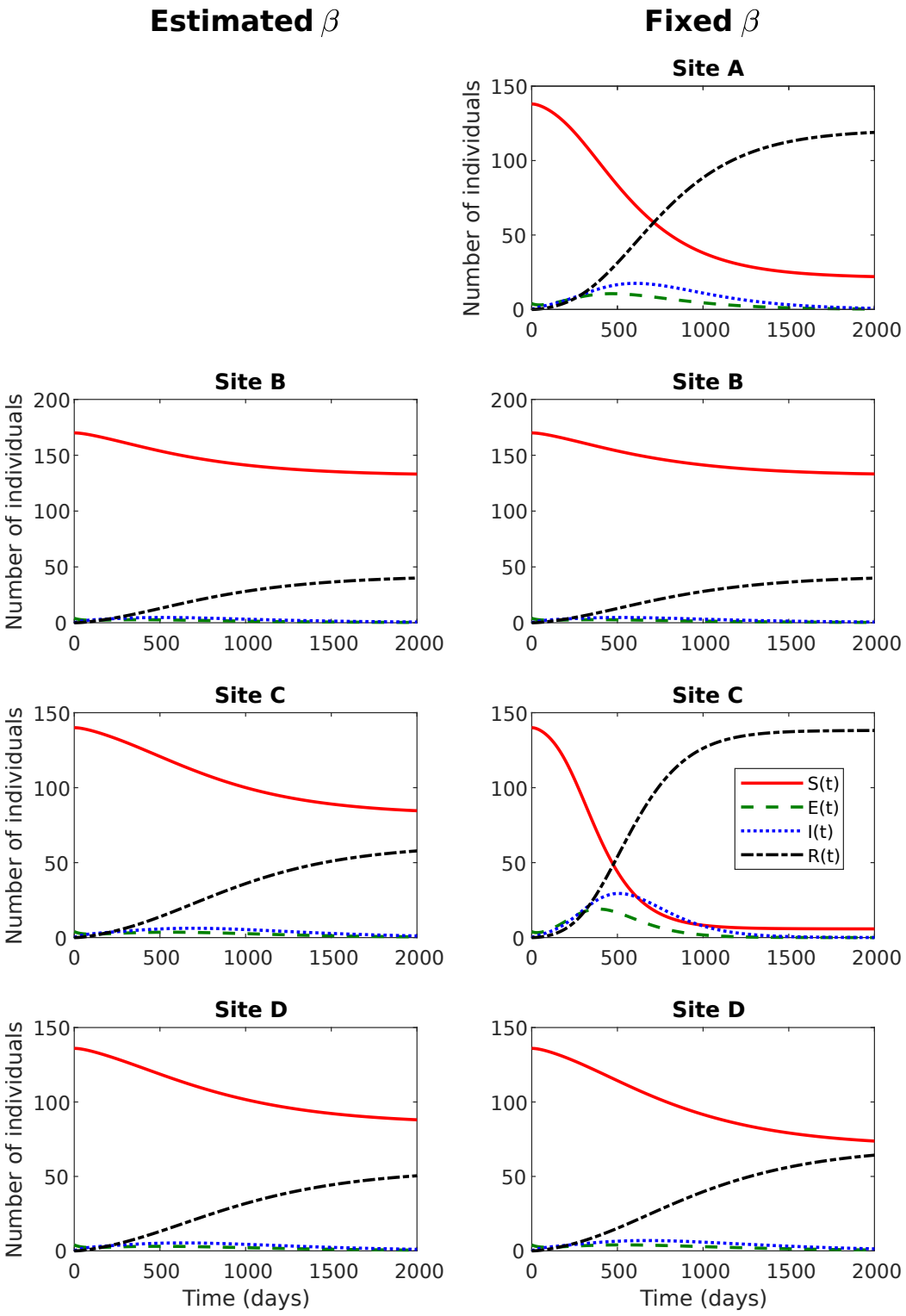



Figure 5.4: **The average number of individuals in each compartment at the end of each day, for simulations on the empirical weighted contact networks.** These values are calculated by averaging over 50 simulations of an epidemic on the 2000 different model networks. Each simulation had a different seed population, chosen by selecting individuals with similar degree ranking as the real-world seed population.

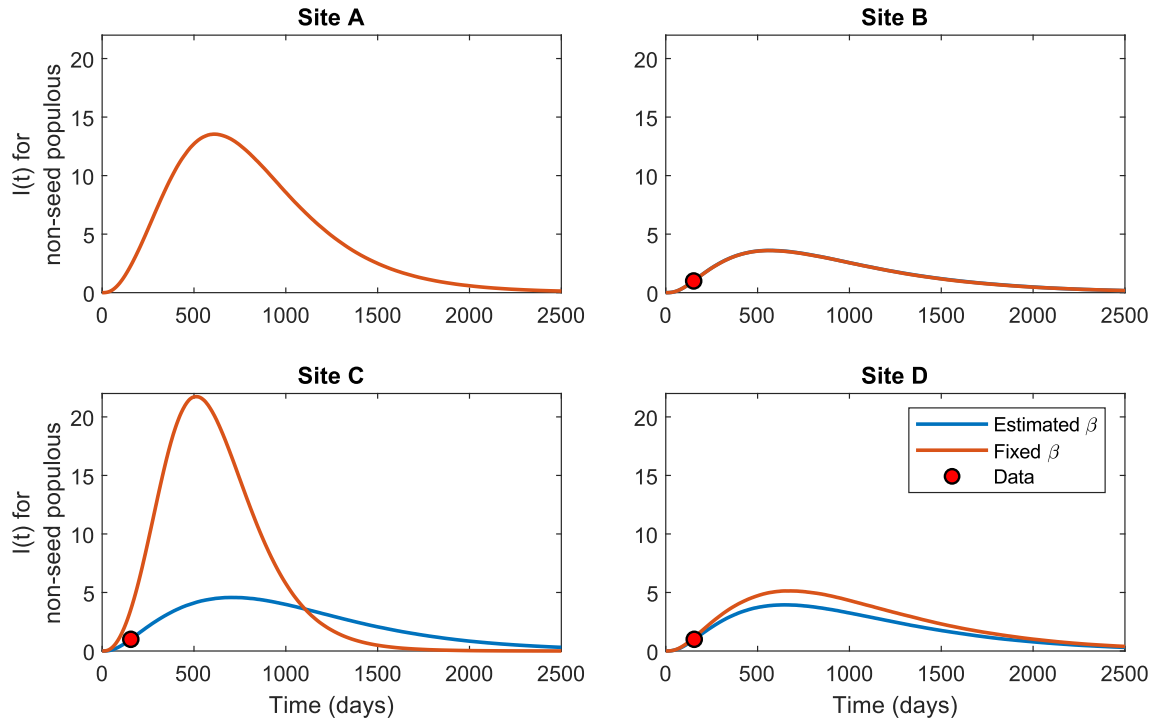



Figure 5.5: **The number of non-collared secondary infectious individuals is well estimated when using the estimated $\beta$.**

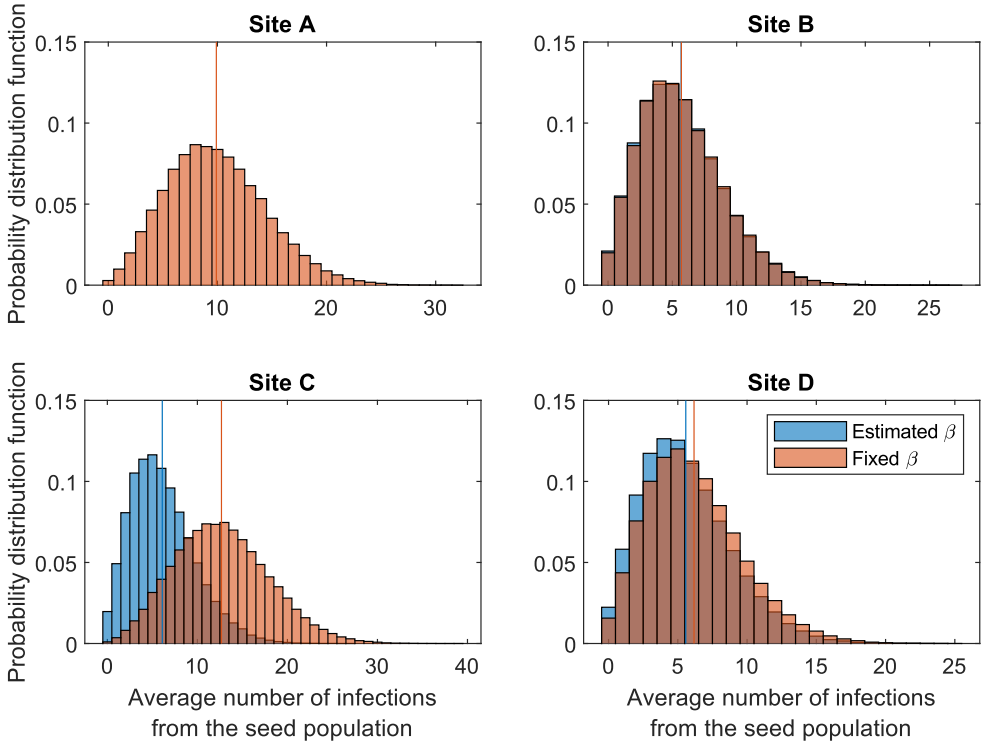



Figure 5.6: **The value of $R_0$ is similar to that of values from the literature.** There are four individuals in each seed population, so we divide the values shown above by four in order to compare with the literature. The average number of secondary infections, that is the estimated $R_0$ value, is shown by the vertical line, and is approximately within the range 1.13 to 2.13 on each site, regardless of which values of $\beta$ is used.

## 5.4 Discussion

We presented a weighted contact network model, built upon networks produced from the possum movement and contact models discussed in Chapter 3. We used this network to simulate bovine tuberculosis epidemics, with an aim to replicating the secondary infections seen in the data set described in Chapter 2.

Moving from a classical SEIR-model to a network-based model requires us to examine how the transmission coefficient $\beta$ is calculated. We estimate values by performing simulations with a range of $\beta$ values, finding the value that produces results closest to the secondary infection data. We also chose a fixed value of $\beta$, so as to overestimate rather than underestimate the number of secondary infections. This gives us, for each model and each site, two values of $\beta$ to use, with the exception of site A, which had no empirical data to use.

We then used the transmission coefficients o simulate epidemic on both network models detailed in Section 5.2. Performing simulations on the model network we find that the average number of secondary infections from a member of the seed population, that is the value of $R_0$, is similar to values in the literature of $R_0$ estimated for TB in possum. Further, we showed that regardless of the transmission coefficient used, the number of non-collared secondary infections is reasonably well-estimated.

In summary, we have developed a stochastic transmission model that, when utilising a spatial network model to describe population structure, allows us to replicate aspects of bovine tuberculosis spread in common brushtail possums that is seen in empirical data.

# Chapter 6

# Concluding remarks

The common brushtail possum (*Trichosurus vulpecula*) is an invasive pest in New Zealand, having been introduced to the country from its native Australian homeland in the 1850s to establish a fur trade, and subsequently released into the wild. It has many damaging effects on the flora and fauna of New Zealand, which is often particularly vulnerable to invasions from abroad. Furthermore, they are also a vector of bovine tuberculosis to cattle which, given the beef and dairy industry's prominence to the New Zealand economy, is of concern to farmers and government alike. Because of these reasons, various organisation, groups and bodies are investigating methods to eradicate the possum population, or if that is not feasible, at least control their numbers.

**Thesis summary**

The research conducted in this thesis aimed to provide insight into possum movement, contact nature and structure, and transmission dynamics in order to aid the efforts currently underway. Our work was motivated by two data sets collected on the contacts between, and TB infection statuses of, possum populations that lived in The North Island.

Our work started with the analysis of these two aforementioned data sets. We cleaned the first data set, removing any daytime or diurnal records, as the model we presented in Chapter 3 only models nocturnal behaviour. Following that, we analysed our cleaned data, showing how it were possible to produce metrics describing the length and importance of the contacts between possum pairs, in addition to network representations. The data sets were used through our work, to inform and verify our models.

Our next step was to develop an individual-based model to describe the movements of, and contacts between, individuals in a population. Individuals move on a square lattice according to a random walk with a given diffusion coefficient, a parameter that describes the rate at which dispersal into an environment occurs. Pairs of individuals can also come into contact when they occupy the same lattice space. By performing simulations of our model for a range of diffusion coefficients and initial distances between dens, we produced measures of the length and importance of these contacts are produced. The probability of two individuals coming into contact during a given night, or the first time a contact occurs during a night where two possums meet are two examples.

Possums typically occupy an area, known as a home-range. We discussed three models for producing distributions of the centres of home-ranges of collared individuals. By then comparing results from model populations to the contact data, we were able to determine a diffusion coefficient that produces populations that are somewhat realistic, and use those diffusion coefficients to estimate home-range sizes. Following that, we built binary and weighted networks from our home-range centre distribution models, and again compared to data. Overall, the first of our home-range models performs best for each site. The implication of our work is that we can produce for contact data a spatial component when the data has no such inherent spatial information component. This could be useful for the planning of future field studies. GPS technology is expensive but, our model shows a possible route such that the technology is not needed.

Due to lethality of bovine tuberculosis to possums, it has been proposed as a possible method for population control. The next part of our work was to build a stochastic transmission model of the spread of this disease through a population. We implemented a Gillespie algorithm to simulate epidemic spread on a network of individuals. After detailing the outputs that can be produced from our model, we investigated what effect changing epidemic parameters, network characteristics including density and edge weight, and the initially infected nodes has on the outcome of a simulation. This was done by simulating epidemics on toy networks. Overall, changing the epidemic parameters, namely the transmission coefficient $\beta$, rate of becoming infectious $\gamma$, and removal rate $\delta$, or the network density has the greatest effect. This effect is seen in both the final size of the epidemic, as well as the number of individuals infected by the initial seed case.

Once our toy network investigation was concluded, we moved forward into presenting a network model upon which we simulate epidemics. These networks are built on the networks produced in Chapter 3. We estimated

values of the transmission coefficient by finding the $\beta$ that gives similar results, and then secondly fixing a value of $\beta$ such that on each s to the secondary infection data. By then simulating epidemics on our model networks, using our estimated $\beta$s, in addition to a larger fixed value, we showed that our contact model networks reproduces aspects of the infection data. The implication of this work is that, by showing our model can recreate disease dynamics, it could be possible to determine how best to exacerbate or reduce an epidemic in a possum population. The former would be useful for population control and for reducing the incurred costs.

**Limitations of the data**

For each of our data sets, that we use to verify our models and are described in Chapter 2, a limitation is present. With the first, that is of contacts between possum pairs, the data we have only records the times and participants of the contacts. As there is a lack of geospatial component, this unfortunately means that we can only infer information about the geographical distribution of the possums, and are unable to know how they actually moved through their environment.

The second data set, detailing secondary TB infections, is limited by its size. Even though the populations of possums on each site number approximately 140 or more, very few secondary infections were recorded. This could have been a result of not finding infected possums, or because of the disease spreading potentially slowly. Whatever the reason, the limited data prevents us from using data points to estimate parameters such as the transmission coefficient $\beta$.

**Limitations of our models and possible future research**

There is definite scope for the expansion and continuation of our work. When considering our individual-based possum model, discussed in Chapter 3, there are certain assumptions that have been made, which lend themselves to further inquiry.

The distribution of denning sites, or home-range centres, where we assumed that they were distributed according to a spatial Poisson point process. However, possums may not be distributed that way. As typically asocial creatures, they may try to den in locations that afford them a large distance from any other possum. Therefore, a distribution method that maximises the pairwise distances might be more appropriate. Or, it may be in fact the opposite. Simultaneous den-sharing between possums does occur [Fairweather et al., 1987],

as may sequential den-sharing [Green and Coleman, 1987]. In those cases, we would desire to cluster the dens in some manner.

This model only models the movement of possums until they first come into contact with one another. When possums come in contact with each other, their behaviour changes, so a simple random walk, where contacts have no affect on the movements of the random walkers, is potentially not realistic for modelling the interactions of possums. One possible avenue that we could explore is the idea of a multi-level model, where the movements of possums before and after their contacts are modelled using a simple random walk as before, but their behaviour whilst in contact is modelled differently. This technique was first introduced by Morales et al. (2004), and has also been used to model possum movements, as in Postlethwaite and Dennis (2013).

The environment in which the possums are moving is assumed geographically uniform, with no boundaries, obstacles, or attractive features including water or food sources. This is of course unrealistic. However, given our lack of information about the terrain of the possums' habitat, this is not a viable avenue of inquiry for this data set, but may be for a different data set.

In addition, our model assumes that all individuals behave in the same way, all the time. This is unrealistic, as male and female possums may behave differently, and seasonal changes will affect behaviour. Further, we make no distinction between non-mating and mating periods, when the behaviour of possums in those two periods will definitely exhibit differences. We also do not taken into account juvenile vs adult behaviour. Juveniles often remain close to their mothers, and then move a significant distance away to establish their own distinct home-range. Our model doesn't reflect that.

The stochastic transmission model that we discussed in Chapter 4 is built upon some assumptions that lend themselves to possible future avenues of research. For example, our model assumes not only that the population size remains constant (if disease-induced fatalities are included), but that there is no non-disease-induced mortality. This may be unrealistic, as juvenile (as well as adult) possums can migrate into and out of the area we are modelling, and possums can die of other things that just bovine tuberculosis. By including births and non-disease-induced deaths, the model could become more realistic.

Furthermore, the assumption that the contact networks are static, where individuals are in contact and

always have been in contact with the same neighbours, may also be unrealistic. By producing random, weighted networks from our contact model, with the edge weights performing as edge probabilities, we could a time-series of networks, representing the contact structure at each time interval. Finally, we assume that the transmission coefficient $\beta$, rate of becoming infectious $\gamma$, and removal rate $\delta$ are identical for each individual. As discussed in Lloyd-Smith et al. (2005), there is likely to be heterogeneity in the population with regards to their susceptibility to a disease, and their infectiousness. A possible solution is to draw parameters from a distribution, rather than have them identical for each individual.

# Bibliography

[Abouelkheir et al., 2017] Abouelkheir, I., Rachik, M., Zakary, O., and Elmouki, I. (2017). A multi-regions sis discrete influenza pandemic model with a travel-blocking vicinity optimal control approach on cells. *Am. J. Comput. Appl. Math*, 7(2):37–45.

[Acedo et al., 2011] Acedo, L., Morano, J.-A., Villanueva, R.-J., Villanueva-Oller, J., and Díez-Domingo, J. (2011). Using random networks to study the dynamics of respiratory syncytial virus (rsv) in the spanish region of valencia. *Mathematical and Computer Modelling*, 54(7-8):1650–1654.

[Ameri and Cooper, 2019] Ameri, K. and Cooper, K. D. (2019). A network-based compartmental model for the spread of whooping cough in nebraska. *AMIA Summits on Translational Science Proceedings*, 2019:388.

[Anderson et al., 2014] Anderson, L. G., White, P. C., Stebbing, P. D., Stentiford, G. D., and Dunn, A. M. (2014). Biosecurity and vector behaviour: evaluating the potential threat posed by anglers and canoeists as pathways for the spread of invasive non-native species and pathogens. *PLoS One*, 9(4):e92788.

[Anderson et al., 1986] Anderson, R. M., Medley, G. F., May, R. M., and Johnson, A. M. (1986). A Preliminary Study of the Transmission Dynamics of the Human Immunodeficiency Virus (HIV), the Causative Agent of AIDS. *IMA Journal of Mathematics Applied in Medicine & Biology*, 3(4):229–263.

[Armstrong, 2014] Armstrong, M. J. (2014). The salvo combat model with a sequential exchange of fire. *Journal of the Operational Research Society*, 65(10):1593–1601.

[Bagler, 2008] Bagler, G. (2008). Analysis of the airport network of india as a complex weighted network. *Physica A: Statistical Mechanics and its Applications*, 387(12):2972–2980.

[Bai et al., 2004] Bai, Y., Thompson, D., and Broersma, K. (2004). Douglas fir and ponderosa pine seed dormancy as regulated by grassland seedbed conditions. *Rangeland Ecology and Management*, 57(6):661–667.

[Barlow, 1993] Barlow, N. (1993). A model for the spread of bovine tb in new zealand possum populations. *Journal of applied ecology*, pages 156–164.

[Barlow, 1991] Barlow, N. D. (1991). A spatially aggregated disease/host model for bovine Tb in New Zealand possum populations. *Journal of Applied Ecology*, 28(3):777–793.

[Barlow, 2000] Barlow, N. D. (2000). Non-linear transmission and simple models for bovine tuberculosis. *Journal of Animal Ecology*, 69(4):703–713.

[Barrat et al., 2004] Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, 101(11):3747–3752.

[Barton et al., 2007] Barton, J., Fowler, S. V., Gianotti, A. F., Winks, C. J., De Beurs, M., Arnold, G. C., and Forrester, G. (2007). Successful biological control of mist flower (ageratina riparia) in new zealand: agent establishment, impact and benefits to the native flora. *Biological Control*, 40(3):370–385.

[Basse et al., 1999] Basse, B., McLennan, J., and Wake, G. (1999). Analysis of the impact of stoats, mustela erminea, on northern brown kiwi, apteryx mantelli, in new zealand. *Wildlife Research*, 26(2):227–237.

[Batcheler and Cowan, 1988] Batcheler, C. and Cowan, P. (1988). *Review of the status of the possum in New Zealand*. Department of Conservation.

[Bichara et al., 2014] Bichara, D., Iggidr, A., and Sallet, G. (2014). Global analysis of multi-strains sis, sir and msir epidemic models. *Journal of Applied Mathematics and Computing*, 44(1):273–292.

[Binny et al., 2016] Binny, R. N., James, A., and Plank, M. J. (2016). Collective cell behaviour with neighbour-dependent proliferation, death and directional bias. *Bulletin of mathematical biology*, 78(11):2277–2301.

[Blackwood and Childs, 2018] Blackwood, J. C. and Childs, L. M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5(1):195–221.

[Blyuss, 2016] Blyuss, K. B. (2016). Mathematical modelling of the dynamics of meningococcal meningitis in africa. In *UK Success Stories in Industrial Mathematics*, pages 221–226. Springer.

[Böckh, 1886] Böckh, R. (1886). Statistisches fahrbuch der stadt berlin, zwölfter jahrgang. *Statistik des Jahres*, 1884:30–31.

[Boguná and Pastor-Satorras, 2002] Boguná, M. and Pastor-Satorras, R. (2002). Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104.

[Brauer, 2008] Brauer, F. (2008). Compartmental models in epidemiology. In *Mathematical epidemiology*, pages 19–79. Springer.

[Britton, 2010] Britton, T. (2010). Stochastic epidemic models: a survey. *Mathematical biosciences*, 225(1):24–35.

[Brockie et al., 1987] Brockie, R., Hearfield, M., White, A., Waddington, D., and Hay, J. (1987). Bovine tuberculosis in a possum from the orongorongo valley, wellington. *New Zealand veterinary journal*, 35(12):201–203.

[Buddenhagen et al., 2017] Buddenhagen, C., Hernandez Nopsa, J.-F., Andersen, K. F., Andrade-Piedra, J., Forbes, G.-A., Kromann, P., Thomas-Sharma, S., Useche, P., and Garrett, K. (2017). Epidemic network analysis for mitigation of invasive pathogens in seed systems: Potato in ecuador. *Phytopathology*, 107(10):1209–1218.

[Byrom et al., 2016] Byrom, A. E., Innes, J., and Binny, R. N. (2016). A review of biodiversity outcomes from possum-focused pest control in new zealand. *Wildlife Research*, 43(3):228–253.

[Campbell et al., 1990] Campbell, D. et al. (1990). Changes in structure and composition of a new zealand lowland forest inhabited by brushtail possums. *Pacific Science*, 44(3):277–296.

[Carey, 2018] Carey, J. (2018). News feature: The race to extinguish insect pests by enlisting their own kind. *Proceedings of the National Academy of Sciences*, 115(31):7839–7843.

[Cator and Van Mieghem, 2012] Cator, E. and Van Mieghem, P. (2012). Second-order mean-field susceptible-infected-susceptible epidemic threshold. *Physical review E*, 85(5):056111.

[Chakrabarti et al., 2008] Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., and Faloutsos, C. (2008). Epidemic thresholds in real networks. *ACM Transactions on Information and System Security*, 10(4):1–26.

[Clout and Ericksen, 2000] Clout, M. and Ericksen, K. (2000). Anatomy of a disastrous success: the brushtail possum as an invasive species. *The Brushtail Possum-Biology, Impact and Management of an Introduced Marsupial*, pages 1–9.

[Codling et al., 2008] Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834.

[Colautti and MacIsaac, 2004] Colautti, R. I. and MacIsaac, H. J. (2004). A neutral terminology to define â€˜invasiveâ€™species. *Diversity and distributions*, 10(2):135–141.

[Conn, 2014] Conn, D. B. (2014). Aquatic invasive species and emerging infectious disease threats: A one health perspective. *Aquatic Invasions*, 9(3).

[Connelly et al., 2007] Connelly, N. A., Oâ€™Neill, C. R., Knuth, B. A., and Brown, T. L. (2007). Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environmental management*, 40(1):105–112.

[Cooke et al., 1999] Cooke, M., Buddle, B., Aldwell, F., McMurray, D., and Alley, M. (1999). The pathogenesis of experimental endo-bronchial mycobacterium bovis infection in brushtail possums (trichosurus vulpecula). *New Zealand Veterinary Journal*, 47(6):187–192.

[Cooke et al., 1995] Cooke, M., Jackson, R., Coleman, J., and Alley, M. (1995). Naturally occurring tuberculosis caused by mycobacterium bovis in brushtail possums (trichosurus vulpecula): Ii. pathology. *New Zealand Veterinary Journal*, 43(7):315–321.

[Cooper and Millener, 1993] Cooper, R. A. and Millener, P. R. (1993). The New Zealand biota: Historical background and new research. *Trends in Ecology & Evolution*, 8(12):429–433.

[Corner et al., 2003a] Corner, L., Buddle, B., and Morris, R. (2003a). Experimental infection of brushtail possums (trichosurus vulpecula) with mycobacterium bovis by conjunctival instillation. *The Veterinary Journal*, 166(2):177–184.

[Corner et al., 2002] Corner, L., Pfeiffer, D., De Lisle, G., Morris, R., and Buddle, B. (2002). Natural transmission of Mycobacterium bovis infection in captive brushtail possums (Trichosurus vulpecula). *New Zealand Veterinary Journal*, 50(4):154–162.

[Corner et al., 2003b] Corner, L., Pfeiffer, D., and Morris, R. (2003b). Social-network analysis of Mycobacterium bovis transmission among captive brushtail possums (Trichosurus vulpecula). *Preventive Veterinary Medicine*, 59(3):147–167.

[Cowan, 2000] Cowan, P. (2000). Biological control of possums: prospects for the future. *The brushtail possum: biology, impact and management of an introduced marsupial. Manaaki Whenua Press, Lincoln, New Zealand*, pages 262–270.

[Cowan and Clout, 2000] Cowan, P. and Clout, M. (2000). Possums on the move: activity patterns, home ranges, and dispersal. *The brushtail possum: biology, impact and management of an introduced marsupial. Manaaki Whenua Press, Lincoln, New Zealand*, pages 24–34.

[Craft et al., 2009] Craft, M. E., Volz, E., Packer, C., and Meyers, L. A. (2009). Distinguishing epidemic waves from disease spillover in a wildlife population. *Proceedings of the Royal Society B: Biological Sciences*, 276(1663):1777–1785.

[Darvey and Staff, 1966] Darvey, I. and Staff, P. (1966). Stochastic approach to first-order chemical reaction kinetics. *The Journal of Chemical Physics*, 44(3):990–997.

[Day et al., 2000] Day, T., O'Connor, C., and Matthews, L. (2000). Possum Social Behaviour. *The Brushtail Possum-Biology, Impact and Management of an Introduced Marsupial*, pages 35–46.

[de Lisle, 1993] de Lisle, G. (1993). Bovine tuberculosis-the new zealand problem. In *Proceedings of the New Zealand Grassland Association*, pages 199–199.

[Department of Conservation, 2004] Department of Conservation (2004). A pest of plague proportions. Accessed on 2020-07-31.

[Department of Conservation, 2020] Department of Conservation (2020). Towards a Predator Free New Zealand.

[Diekmann et al., 1990] Diekmann, O., Heesterbeek, J. A. P., and Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382.

[Dietz, 1993] Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, 2(1):23–41.

[Drewe et al., 2011] Drewe, J. A., Eames, K. T., Madden, J. R., and Pearce, G. P. (2011). Integrating contact network structure into tuberculosis epidemiology in meerkats in south africa: Implications for control. *Preventive Veterinary Medicine*, 101(1):113 – 120.

[Effler et al., 1996] Effler, S. W., Brooks, C. M., Whitehead, K., Wagner, B., Doerr, S. M., Perkins, M., Siegfried, C. A., Walrath, L., and Canale, R. P. (1996). Impact of zebra mussel invasion on river water quality. *Water Environment Research*, 68(2):205–214.

[Efford, 2004] Efford, M. G. (2004). Density estimation in live-trapping studies. *Oikos*, 106(3):598–610.

[Efford and Fewster, 2013] Efford, M. G. and Fewster, R. M. (2013). Estimating population size by spatially explicit capture–recapture. *Oikos*, 122(6):918–928.

[Efford et al., 2005] Efford, M. G., Warburton, B., Coleman, M. C., and Barker, R. J. (2005). A field test of two methods for density estimation. *Wildlife Society Bulletin*, 33(2):731–738.

[Ehrhardt et al., 2019] Ehrhardt, M., Gašper, J., and Kilianová, S. (2019). Sir-based mathematical modeling of infectious diseases with vaccination and waning immunity. *Journal of Computational Science*, 37:101027.

[Ekchian, 1982] Ekchian, L. K. (1982). An overview of lanchaster-type combat models for modern warfare scenarios.

[Ekdahl et al., 1970] Ekdahl, M., Smith, B., and Money, D. (1970). Letters to the editor. *New Zealand Veterinary Journal*, 18(3):44–45.

[Erban and Chapman, 2009] Erban, R. and Chapman, S. J. (2009). Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. *Physical biology*, 6(4):046001.

[Estrada et al., 2016] Estrada, E., Meloni, S., Sheerin, M., and Moreno, Y. (2016). Epidemic spreading in random rectangular networks. *Physical Review E*, 94(5):052316.

[Fairweather et al., 1987] Fairweather, A., Brockie, R., and Ward, G. (1987). Possums (trichosurus vulpecula) sharing dens: A potential infection route for bovine tuberculosis. *New Zealand Veterinary Journal*, 35(1-2):15–16.

[Ferguson et al., 2001] Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001). The foot-and-mouth epidemic in great britain: pattern of spread and impact of interventions. *Science*, 292(5519):1155–1160.

[Ferguson et al., 2003] Ferguson, N. M., Keeling, M. J., Edmunds, W. J., Gani, R., Grenfell, B. T., Anderson, R. M., and Leach, S. (2003). Planning for smallpox outbreaks. *Nature*, 425(6959):681–685.

[Fine, 1993] Fine, P. E. (1993). Herd immunity: history, theory, practice. *Epidemiologic reviews*, 15(2):265–302.

[Fletcher, 2017] Fletcher, S. J. (2017). Chapter 3 - univariate distribution theory. In Fletcher, S. J., editor, *Data Assimilation for the Geosciences*, pages 29 – 124. Elsevier.

[Frias-Martinez et al., 2011] Frias-Martinez, E., Williamson, G., and Frias-Martinez, V. (2011). An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE third*

*international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, pages 57–64. IEEE.

[Gail and Boone, 1970] Gail, M. H. and Boone, C. W. (1970). The locomotion of mouse fibroblasts in tissue culture. *Biophysical journal*, 10(10):980.

[Gallardo and Aldridge, 2013] Gallardo, B. and Aldridge, D. C. (2013). The â€˜dirty dozenâ€™: socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in great britain and ireland. *Journal of Applied Ecology*, 50(3):757–766.

[Getz et al., 2019] Getz, W. M., Salter, R., and Mgbara, W. (2019). Adequacy of seir models when epidemics have spatial structure: Ebola in sierra leone. *Philosophical Transactions of the Royal Society B*, 374(1775):20180282.

[Ghani et al., 1998] Ghani, A. C., Donnelly, C. A., and Garnett, G. P. (1998). Sampling Biases and Missing Data in Explorations of Sexual Partner Networks for. *Statistics in Medicine*, 2097(April 1997):2079–2097.

[Gillespie, 1977] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361.

[Gilligan et al., 1997] Gilligan, C. A., Gubbins, S., and Simons, S. A. (1997). Analysis and fitting of an sir model with host response to infection load for a plant disease. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1351):353–364.

[González-Parra et al., 2010] González-Parra, G., Acedo, L., Micó, R.-J. V., and Arenas, A. J. (2010). Modeling the social obesity epidemic with stochastic networks. *Physica A: Statistical Mechanics and its Applications*, 389(17):3692–3701.

[González-Parra et al., 2015] González-Parra, G., Villanueva, R.-J., Ruiz-Baragaño, J., and Moraño, J.-A. (2015). Modelling influenza a (h1n1) 2009 epidemics using a random network in a distributed computing environment. *Acta tropica*, 143:29–35.

[Green and Rohan, 2012] Green, W. and Rohan, M. (2012). Opposition to aerial 1080 poisoning for control of invasive mammals in new zealand: risk perceptions and agency responses. *Journal of the Royal Society of New Zealand*, 42(3):185–213.

[Green and Coleman, 1987] Green, W. Q. and Coleman, J. (1987). Den sites of possums, trichosurus-vulpecula, and frequency of use in mixed hardwood forest in westland, new-zealand. *Wildlife Research*, 14(3):285–292.

[Hackenberger, 2019] Hackenberger, B. K. (2019). M for measles, m for math, m for mod... *Croatian medical journal*, 60(5):463.

[Hailegiorgis and Crooks, 2012] Hailegiorgis, A. and Crooks, A. T. (2012). Agent-based modeling for humanitarian issues: disease and refugee camps. In *The Computational Social Science Society of America Conference, Santa Fe, NM.*

[Halwart, 1994] Halwart, M. (1994). The golden apple snail pomacea canaliculata in asian rice farming systems: present impact and future threat. *International Journal of Pest Management*, 40(2):199–206.

[Hamede et al., 2009] Hamede, R. K., Bashford, J., McCallum, H., and Jones, M. (2009). Contact networks in a wild Tasmanian devil (Sarcophilus harrisii) population: Using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecology Letters*, 12(11):1147–1157.

[Haydon et al., 2008] Haydon, D. T., Morales, J. M., Yott, A., Jenkins, D. A., Rosatte, R., and Fryxell, J. M. (2008). Socially informed random walks: incorporating group dynamics into models of population spread and growth. *Proceedings of the Royal Society B: Biological Sciences*, 275(1638):1101–1109.

[Heffernan et al., 2005] Heffernan, J. M., Smith, R. J., and Wahl, L. M. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293.

[Higham et al., 2008] Higham, D. J., Rašajski, M., and Pržulj, N. (2008). Fitting a geometric graph to a protein-protein interaction network. *Bioinformatics*, 24(8):1093–1099.

[Hill and Häder, 1997] Hill, N. and Häder, D.-P. (1997). A biased random walk model for the trajectories of swimming micro-organisms. *Journal of Theoretical Biology*, 186(4):503–526.

[Horan and Lupi, 2005] Horan, R. D. and Lupi, F. (2005). Tradeable risk permits to prevent future introductions of invasive alien species into the great lakes. *Ecological Economics*, 52(3):289–304.

[Howe, 2018] Howe, S. (2018). Rabbit virus working as expected â€“ the science behind RHDV1-K5.

[Innes et al., 2010] Innes, J., Kelly, D., Overton, J. M., and Gillies, C. (2010). Predation and other factors currently limiting new zealand forest birds. *New Zealand Journal of Ecology*, 34(1):86.

[Jackson et al., 1995] Jackson, R., Cooke, M., Coleman, J., Morris, R., de Lisle, G., and Yates, G. (1995). Naturally occurring tuberculosis caused by Mycobacterium bovis in brushtail possums (Trichosurus vulpecula): III. Routes of infection and excretion. *New Zealand Veterinary Journal*, 43(7):322–327. PMID: 16031873.

[James et al., 2017] James, A., McLeod, J. C., Rouco, C., Richardson, K. S., and Tompkins, D. M. (2017). Spatial utilization predicts animal social contact networks are not scale-free. *Royal Society Open Science*, 4(12).

[Jarman and Ballschmiter, 2012] Jarman, W. M. and Ballschmiter, K. (2012). From coal to ddt: the history of the development of the pesticide ddt from synthetic dyes till silent spring. *Endeavour*, 36(4):131–142.

[Jarvis et al., 2006] Jarvis, P., Fowler, S., Paynter, Q., and Syrett, P. (2006). Predicting the economic benefits and costs of introducing new biological control agents for scotch broom cytisus scoparius into new zealand. *Biological Control*, 39(2):135–146.

[Ji et al., 2005] Ji, W., White, P. C., and Clout, M. N. (2005). Contact rates between possums revealed by proximity data loggers. *Journal of Applied Ecology*, 42(3):595–604.

[Johnson and MacKay, 2011] Johnson, I. R. and MacKay, N. J. (2011). Lanchester models and the battle of britain. *Naval Research Logistics (NRL)*, 58(3):210–222.

[Johnsson et al., 1996] Johnsson, A., Karlsson, C., Chapman, D., Braseth, J., and Iversen, T.-H. (1996). Dynamics of root growth in microgravity. *Journal of biotechnology*, 47(2-3):155–165.

[Jolly, 1993] Jolly, S. (1993). Biological control of possums. *New Zealand journal of zoology*, 20(4):335–339.

[Julia et al., 2007] Julia, R., Holland, D. W., and Guenthner, J. (2007). Assessing the economic impact of invasive species: the case of yellow starthistle (centaurea solsitialis l.) in the rangelands of idaho, usa. *Journal of Environmental Management*, 85(4):876–882.

[Kamina et al., 2019] Kamina, K. M., Mwalili, S., and Wanjoya, A. (2019). The modeling of a stochastic sir model for hiv/aids epidemic using gillespieâ€™s algorithm. *Int. J. Data Sci. Anal*, 5(6):117.

[Karaman et al., 2016] Karaman, M. M., Sui, Y., Wang, H., Magin, R. L., Li, Y., and Zhou, X. J. (2016). Differentiating low-and high-grade pediatric brain tumors using a continuous-time random-walk diffusion model at high b-values. *Magnetic resonance in medicine*, 76(4):1149–1157.

[Kareiva and Shigesada, 1983] Kareiva, P. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia*, 56(2-3):234–238.

[Keeling and Eames, 2005] Keeling, M. J. and Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307.

[Keeling et al., 2001] Keeling, M. J., Woolhouse, M. E., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J., and Grenfell, B. T. (2001). Dynamics of the 2001 uk foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817.

[Keller et al., 2011] Keller, R. P., Geist, J., Jeschke, J. M., and Kühn, I. (2011). Invasive species in europe: ecology, status, and policy. *Environmental Sciences Europe*, 23(1):23.

[Kenah and Robins, 2007] Kenah, E. and Robins, J. M. (2007). Second look at the spread of epidemics on networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 76(3):1–12.

[Kenniche and Ravelomananana, 2010] Kenniche, H. and Ravelomananana, V. (2010). Random geometric graphs as model of wireless sensor networks. *2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010*, 4:103–107.

[Kephart and White, 1992] Kephart, J. O. and White, S. R. (1992). Directed-graph epidemiological models of computer viruses. In *Computation: the micro and the macro view*, pages 71–102. World Scientific.

[Kermack and McKendrick, 1927] Kermack, W. O. and McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721.

[Kibona and Yang, 2017] Kibona, I. E. and Yang, C. (2017). Sir model of spread of zika virus infections: Zikv linked to microcephaly simulations. *Health*, 9(8):1190–1210.

[King and Moody, 1982] King, C. and Moody, J. (1982). The biology of the stoat (mustela erminea) in the national parks of new zealand i. general introduction. *New Zealand journal of zoology*, 9(1):49–55.

[King, 1990] King, C. M. (1990). *The Handbook of New Zealand Mammals*. Oxford University Press.

[Kohlmeier and Ebenhöh, 1995] Kohlmeier, C. and Ebenhöh, W. (1995). The stabilizing role of cannibalism in a predator-prey system. *Bulletin of Mathematical Biology*, 57(3):401–411.

[Kossinets, 2006] Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, 28(3):247–268.

[Kucharski et al., 2015] Kucharski, A. J., Camacho, A., Flasche, S., Glover, R. E., Edmunds, W. J., and Funk, S. (2015). Measuring the impact of ebola control measures in sierra leone. *Proceedings of the National Academy of Sciences*, 112(46):14366–14371.

[Laager et al., 2018] Laager, M., Mbilo, C., Madaye, E. A., Naminou, A., Léchenne, M., Tschopp, A., Smieszek, T., Zinsstag, J., Chitnis, N., et al. (2018). The importance of dog population contact network structures in rabies transmission. *PLoS neglected tropical diseases*, 12(8):e0006680.

[Landcare Research, 2000] Landcare Research (2000). Possums and TB. Accessed on 2017-04-16.

[Landcare Research, 2002] Landcare Research (2002). Effects of Possums on Native Animals. Accessed on 2017-04-16.

[Lang and De Sterck, 2014] Lang, J. C. and De Sterck, H. (2014). The arab spring: A simple compartmental model for the dynamics of a revolution. *Mathematical Social Sciences*, 69:12–21.

[Le Maitre et al., 2011] Le Maitre, D., De Lange, W., Richardson, D., Wise, R., and Van Wilgen, B. (2011). The economic consequences of the environmental impacts of alien plant invasions in south africa. *Biological invasions: Economic and environmental costs of alien plant, animal, and microbe species*, pages 295–323.

[Lekone and Finkenstädt, 2006] Lekone, P. E. and Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic seir model with control intervention: Ebola as a case study. *Biometrics*, 62(4):1170–1177.

[Levine and D'Antonio, 2003] Levine, J. M. and D'Antonio, C. M. (2003). Forecasting biological invasions with increasing international trade. *Conservation Biology*, 17(1):322–326.

[Lin et al., 2010] Lin, P.-C., Wu, B., and Watada, J. (2010). Kolmogorov-smirnov two sample test with continuous fuzzy data. In *Integrated Uncertainty Management and Applications*, pages 175–186. Springer.

[Lindenberg et al., 1980] Lindenberg, K., Seshadri, V., Shuler, K., and Weiss, G. H. (1980). Lattice random walks for sets of random walkers. first passage times. *Journal of Statistical Physics*, 23(1):11–25.

[Liu et al., 1999] Liu, J. H., Wilson, M. S., McClure, J., and Higgins, T. R. (1999). Social identity and the perception of history: Cultural representations of aotearoa/new zealand. *European Journal of Social Psychology*, 29(8):1021–1047.

[Lloyd-Smith et al., 2005] Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. (2005). Super-spreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359.

[Lovely and Dahlquist, 1975] Lovely, P. S. and Dahlquist, F. (1975). Statistical measures of bacterial motility and chemotaxis. *Journal of theoretical biology*, 50(2):477–496.

[Lowe, 2008] Lowe, D. J. (2008). Polynesian settlement of new zealand and the impacts of volcanism on early maori society: An update. *New Zealand Society of Soil Science*, pages 142–147.

[Ludyanskiy et al., 1993] Ludyanskiy, M. L., McDonald, D., and MacNeill, D. (1993). Impact of the zebra mussel, a bivalve invader: Dreissena polymorpha is rapidly colonizing hard surfaces throughout waterways of the united states and canada. *Bioscience*, 43(8):533–544.

[Mack et al., 2000] Mack, R. N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., and Bazzaz, F. A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological applications*, 10(3):689–710.

[MacLennan, 1984] MacLennan, D. (1984). The feeding behaviour and activity patterns of the brushtail possum, trichosurus vulpecula, in an open eucalypt woodland in southeast queensland. *Possums and gliders*, pages 155–161.

[Malik et al., 2017] Malik, H. A. M., Mahesar, A. W., Abid, F., Waqas, A., and Wahiddin, M. R. (2017). Two-mode network modeling and analysis of dengue epidemic behavior in gombak, malaysia. *Applied Mathematical Modelling*, 43:207–220.

[Manaaki Whenua - Landcare Research, 2016] Manaaki Whenua - Landcare Research (2016). Rabbit biocontrol in New Zealand RHDV1 K5.

[Manaaki Whenua - Landcare Research, 2018] Manaaki Whenua - Landcare Research (2018). More about RHDV1 K5.

[May and Anderson, 1987] May, R. M. and Anderson, R. M. (1987). Transmission dynamics of HIV infection. *Nature*, 326(6109):137–142.

[May and Anderson, 1988] May, R. M. and Anderson, R. M. (1988). The transmission dynamics of human immunodeficiency virus (hiv). *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 321(1207):565–607.

[McCallum et al., 2001] McCallum, H., Barlow, N., and Hone, J. (2001). How should pathogen transmission be modelled? *Trends in ecology & evolution*, 16(6):295–300.

[McLennan, 2006] McLennan, J. (2006). Strategies to reduce predation on bird populations. In *Biological invasions in New Zealand*, pages 371–387. Springer.

[McLennan et al., 1996] McLennan, J., Potter, M., Robertson, H., Wake, G., Colbourne, R., Dew, L., Joyce, L., McCann, A., Miles, J., Miller, P., et al. (1996). Role of predation in the decline of kiwi, apteryx spp., in new zealand. *New Zealand Journal of Ecology*, pages 27–35.

[Meeus, 1998] Meeus, J. (1998). *Astronomical algorithms*. Willmann-Bell, Richmond, Va, 2nd edition.

[Menkhorst et al., 2004] Menkhorst, P., Knight, F., et al. (2004). *A field guide to the mammals of Australia*, volume 2. Oxford University Press Melbourne.

[Meyerson and Mooney, 2007] Meyerson, L. A. and Mooney, H. A. (2007). Invasive alien species in an era of globalization. *Frontiers in Ecology and the Environment*, pages 199–208.

[Michael Hardie Boys, Governor-General, 2005] Michael Hardie Boys, Governor-General (2005). Biosecurity (National Bovine Tuberculosis Pest Management Strategy) Order 1998 (SR 1998/179) (as at 22 December 2005) Contents â€“ New Zealand Legislation.

[Michael Hardie Boys, Governor-General, 2016] Michael Hardie Boys, Governor-General (2016). Biosecurity (National Bovine Tuberculosis Pest Management Plan) Order 1998 (SR 1998/179) (as at 1 August 2016) Contents â€“ New Zealand Legislation.

[Milledge et al., 2016] Milledge, J. J., Nielsen, B. V., and Bailey, D. (2016). High-value products from macroalgae: the potential uses of the invasive brown seaweed, sargassum muticum. *Reviews in Environmental Science and Bio/Technology*, 15(1):67–88.

[Mills et al., 2004] Mills, C. E., Robins, J. M., and Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature*, 432(7019):904–906.

[Minchin et al., 2002] Minchin, D., Lucy, F., and Sullivan, M. (2002). Zebra mussel: impacts and spread. In *Invasive aquatic species of Europe. Distribution, impacts and management*, pages 135–146. Springer.

[Ministry for Primary Industries, 2019] Ministry for Primary Industries (2019). Situation and Outlook for Primary Industries December 2019.

[Montroll, 1964] Montroll, E. W. (1964). Random walks on lattices. *Stochastic processes in mathematical physics and engineering. American Mathematical Society.*, pages 193–220.

[Montroll and Weiss, 1965] Montroll, h. W. and Weiss, G. H. (1965). Random walks on lattices. ii. *Journal of Mathematical Physics*, 6(2):167–181.

[Mooney et al., 2005] Mooney, H. A. et al. (2005). Invasive alien species: the nature of the problem. *Scope-Scientific Committee on Problems of the Environment International Council of Scientific Unions*, 63:1.

[Morales et al., 2004] Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.

[Morgan and Hickling, 2000] Morgan, D. and Hickling, G. (2000). Techniques used for poisoning possums. *The brushtail possum: biology, impact and management of an introduced marsupial. Manaaki Whenua Press, Lincoln, New Zealand*, pages 143–153.

[Morris and Pfeiffer, 1995] Morris, R. and Pfeiffer, D. (1995). Directions and issues in bovine tuberculosis epidemiology and control in New Zealand. *New Zealand Veterinary Journal*, 43(7):256–265.

[Morriss et al., 2020] Morriss, G. A., Parkes, J. P., and Nugent, G. (2020). Effects of aerial 1080 operations on deer populations in new zealand. *New Zealand Journal of Ecology*, 44(2):3417.

[Naylor, 1996] Naylor, R. (1996). Invasions in agriculture: assessing the cost of the golden apple snail in asia. *Ambio*, pages 443–448.

[Newman, 2001] Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409.

[Newman, 2002] Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 66(1):1–11.

[Nielsen et al., 2007] Nielsen, L. R., van den Borne, B., and van Schaik, G. (2007). Salmonella dublin infection in young dairy calves: transmission parameters estimated from field data and an sir-model. *Preventive Veterinary Medicine*, 79(1):46–58.

[Nishiura et al., 2011] Nishiura, H., Chowell, G., and Castillo-Chavez, C. (2011). Did modeling overestimate the transmission potential of pandemic (H1N1-2009)? sample size estimation for post-epidemic seroepidemiological studies. *PLoS ONE*, 6(3).

[Nugent, 1995] Nugent, G. (1995). Effects of possums on the native flora. *Possums as conservation pests*, 29:5.

[Nugent, 2011] Nugent, G. (2011). Maintenance, spillover and spillback transmission of bovine tuberculosis in multi-host wildlife complexes: a new zealand case study. *Veterinary microbiology*, 151(1-2):34–42.

[Nugent et al., 2015] Nugent, G., Buddle, B. M., and Knowles, G. (2015). Epidemiology and control of Mycobacterium bovis infection in brushtail possums (Trichosurus vulpecula), the primary wildlife host of bovine tuberculosis in New Zealand. *New Zealand Veterinary Journal*, 63(0):28–41.

[O'Hara, 2006] O'Hara, P. (2006). The illegal introduction of rabbit haemorrhagic disease virus in new zealand. *Revue scientifique et technique (International Office of Epizootics)*, 25(1):119.

[Omic, 2010] Omic, J. (2010). Epidemics in networks: modeling, optimization and security games.

[Onnela et al., 2007] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., De Menezes, M. A., Kaski, K., Barabási, A.-L., and Kertész, J. (2007). Analysis of a large-scale weighted network of one-to-one human communication. *New journal of physics*, 9(6):179.

[OSPRI New Zealand (OSPRI), 2019] OSPRI New Zealand (OSPRI) (2019). Annual Report 2018/19.

[Otterstatter and Thomson, 2007] Otterstatter, M. C. and Thomson, J. D. (2007). Contact networks and transmission of an intestinal pathogen in bumble bee (bombus impatiens) colonies. *Oecologia*, 154(2):411–421.

[Parkes et al., 2002] Parkes, J., Norbury, G., Heyward, R., and Sullivan, G. (2002). Epidemiology of rabbit haemorrhagic disease (rhd) in the south island, new zealand, 1997–2001. *Wildlife Research*, 29(6):543–555.

[Pastor-Satorras and Vespignani, 2002] Pastor-Satorras, R. and Vespignani, A. (2002). Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3):035108.

[Paterson and Morris, 1995] Paterson, B. and Morris, R. (1995). Interactions between beef cattle and simulated tuberculous possums on pasture. *New Zealand Veterinary Journal*, 43(7):289–293.

[Paupy et al., 2012] Paupy, C., Kassa Kassa, F., Caron, M., Nkoghé, D., and Leroy, E. M. (2012). A chikungunya outbreak associated with the vector aedes albopictus in remote villages of gabon. *Vector-borne and zoonotic diseases*, 12(2):167–169.

[Pejchar and Mooney, 2009] Pejchar, L. and Mooney, H. A. (2009). Invasive species, ecosystem services and human well-being. *Trends in ecology & evolution*, 24(9):497–504.

[Perrings et al., 2005] Perrings, C., Dehnen-Schmutz, K., Touza, J., and Williamson, M. (2005). How to manage biological invasions under globalization. *Trends in ecology & evolution*, 20(5):212–215.

[Pfeiffer et al., 1995] Pfeiffer, D., Hickling, G., Morris, R., Patterson, K., Ryan, T., and Crews, K. (1995). The epidemiology of mycobacterium bovis infection in brushtail possums (trichosurus vulpecula kerr) in the hauhungaroa ranges, new zealand. *New Zealand Veterinary Journal*, 43(7):272–280.

[Pimentel, 2002] Pimentel, D. (2002). Introduction: non-native species in the world. *Biological invasions: economic and environmental costs of alien plant, animal, and microbe species*, pages 3–8.

[Pimentel, 2011] Pimentel, D. (2011). Environmental and economic costs associated with alien invasive species in the united states. *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species. CRC Press, Boca Raton*, pages 411–430.

[Pimentel et al., 2005] Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the united states. *Ecological economics*, 52(3):273–288.

[Pinteus et al., 2020] Pinteus, S., Lemos, M. F., Alves, C., Silva, J., and Pedrosa, R. (2020). The marine invasive seaweeds asparagopsis armata and sargassum muticum as targets for greener antifouling solutions. *Science of The Total Environment*, 750:141372.

[Pitchford and Brindley, 1998] Pitchford, J. and Brindley, J. (1998). Intratrophic predation in simple predator–prey models. *Bulletin of Mathematical Biology*, 60(5):937–953.

[Porphyre et al., 2008] Porphyre, T., Stevenson, M., Jackson, R., and McKenzie, J. (2008). Influence of contact heterogeneity on tb reproduction ratio r _0 in a free-living brushtail possum trichosurus vulpecula population. *Veterinary Research*, 39(3):1.

[Postlethwaite and Dennis, 2013] Postlethwaite, C. M. and Dennis, T. E. (2013). Effects of temporal resolution on an inferential model of animal movement. *PLoS One*, 8(5).

[Pracy, 1974] Pracy, L. (1974). Introduction and liberation of the opossum (trichosurus vulpeca) into new zealand. government printer, wellington. *New Zealand Forest Service Information Series No*, 45.

[Preciado and Jadbabaie, 2009] Preciado, V. M. and Jadbabaie, A. (2009). Spectral analysis of virus spreading in random geometric networks. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 4802–4807. IEEE.

[Ramsey et al., 2009] Ramsey, D., Aldwell, F., Cross, M., de Lisle, G., and Buddle, B. (2009). Protection of free-living and captive possums against pulmonary challenge with Mycobacterium bovis following oral BCG vaccination. *Tuberculosis*, 89(2):163–168.

[Ramsey and Cowan, 2003] Ramsey, D. and Cowan, P. (2003). Mortality rate and movements of brushtail possums with clinical tuberculosis (mycobacterium bovis infection). *New Zealand Veterinary Journal*, 51(4):179–185.

[Ramsey and Efford, 2010] Ramsey, D. S. and Efford, M. G. (2010). Management of bovine tuberculosis in brushtail possums in New Zealand: Predictions from a spatially explicit, individual-based model. *Journal of Applied Ecology*, 47(4):911–919.

[Rausanu and Grosan, 2010] Rausanu, S. and Grosan, C. (2010). Social Networks for Epidemic Spreading : A Case Study. *Studia Universitatis Babes-Bolyai, Informatica*, 55(1):71–86.

[Reich and Gardner, 2014] Reich, B. J. and Gardner, B. (2014). A spatial capture-recapture model for territorial species. *Environmetrics*, 25(8):630–637.

[Reynolds et al., 2017] Reynolds, A., Santini, G., Chelazzi, G., and Focardi, S. (2017). The weierstrassian movement patterns of snails. *Royal Society open science*, 4(6):160941.

[Reynolds et al., 2014] Reynolds, J. J., Torremorell, M., and Craft, M. E. (2014). Mathematical modeling of influenza a virus dynamics within swine farms and the effects of vaccination. *PloS one*, 9(8):e106177.

[Richardson, 2016] Richardson, K. S. (2016). *An Australian abroad: the secret life of the brushtail possum (Trichosurus vulpecula): a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Veterinary Science at Massey University, Manawatu, New Zealand.* PhD thesis, Massey University.

[Richardson et al., 2017] Richardson, K. S., Rouco, C., Jewell, C., French, N. P., Buddle, B. M., and Tompkins, D. M. (2017). Investigating brushtail possum (Trichosurus vulpecula) home-range size determinants in a New Zealand native forest. *Wildlife Research*, 44(4):316–323.

[Ríos-Uzeda et al., 2019] Ríos-Uzeda, B., Brigatti, E., and Vieira, M. V. (2019). Lévy like patterns in the small-scale movements of marsupials in an unfamiliar and risky environment. *Scientific reports*, 9(1):1–12.

[Roberts, 1992] Roberts, M. (1992). The dynamics and control of bovine tuberculosis in possums. *Mathematical Medicine and Biology: A Journal of the IMA*, 9(1):19–28.

[Roberts, 1996] Roberts, M. (1996). The dynamics of bovine tuberculosis in possum populations, and its eradication or control by culling or vaccination. *Journal of Animal Ecology*, 65(4):451–464.

[Robineau et al., 2017] Robineau, O., Velter, A., Barin, F., and Boelle, P.-Y. (2017). Hiv transmission and pre-exposure prophylaxis in a high risk msm population: A simulation study of location-based selection of sexual partners. *PloS one*, 12(11):e0189002.

[Rolls et al., 2015] Rolls, D. A., Wang, P., McBryde, E., Pattison, P., and Robins, G. (2015). A simulation study comparing epidemic dynamics on exponential random graph and edge-triangle configuration type contact network models. *PloS one*, 10(11):e0142181.

[Rorres et al., 2011] Rorres, C., Pelletier, S. T., and Smith, G. (2011). Stochastic modeling of animal epidemics using data collected over three different spatial scales. *Epidemics*, 3(2):61–70.

[Rouco et al., 2018] Rouco, C., Jewell, C., Richardson, K., French, N., Buddle, B. M., and Tompkins, D. (2018). Brushtail possum (Trichosurus vulpecula) social interactions and their implications for bovine tuberculosis epidemiology. *Behaviour*, 155(7-9):621–637.

[Ryan et al., 2006] Ryan, T., Livingstone, P., Ramsey, D., De Lisle, G., Nugent, G., Collins, D., and Buddle, B. (2006). Advances in understanding disease epidemiology and implications for control and eradication of tuberculosis in livestock: the experience from new zealand. *Veterinary microbiology*, 112(2-4):211–219.

[Searle et al., 2016] Searle, C. L., Cortez, M. H., Hunsberger, K. K., Grippi, D. C., Oleksy, I. A., Shaw, C. L., de la Serna, S. B., Lash, C. L., Dhir, K. L., and Duffy, M. A. (2016). Population density, not host competence, drives patterns of disease in an invaded community. *The American Naturalist*, 188(5):554–566.

[Sharpe and Garcelon, 2005] Sharpe, P. B. and Garcelon, D. K. (2005). Restoring and monitoring bald eagles in southern california: The legacy of ddt. In *Proceedings of the Sixth California Islands Symposium*, page 323. Institute for Wildlife Studies: Arcata, CA, USA.

[Sidhu et al., 2002] Sidhu, K., Mate, K., Harris, M., Molinia, F., Duckworth, J., Cowan, P., and Rodger, J. (2002). Immunocontraceptive Biotechnology as Respite to Environmental Degradation Caused by the Pest Species: a lesson from brushtail possum overabundance in New Zealand. In *Environmental Biotechnology*, pages 345–362. APH Publishing Corporation.

[Sims et al., 2014] Sims, D. W., Reynolds, A. M., Humphries, N. E., Southall, E. J., Wearmouth, V. J., Metcalfe, B., and Twitchett, R. J. (2014). Hierarchical random walks in trace fossils and the origin of optimal search behavior. *Proceedings of the National Academy of Sciences*, 111(30):11073–11078.

[Smith and Moody, 2013] Smith, J. a. and Moody, J. (2013). Structural effects of network sampling coverage I: Nodes missing at random. *Social Networks*, 35(4):652–668.

[Stehlé et al., 2011] Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Régis, C., Pinton, J.-F., Khanafer, N., Van den Broeck, W., et al. (2011). Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9(1):87.

[Takaguchi et al., 2013] Takaguchi, T., Masuda, N., and Holme, P. (2013). Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics. *PloS one*, 8(7):e68629.

[Taugbøl et al., 1993] Taugbøl, T., Skurdal, J., and Håstein, T. (1993). Crayfish plague and management strategies in norway. *Biological Conservation*, 63(1):75–82.

[Thompson, 2016] Thompson, K. M. (2016). Evolution and use of dynamic transmission models for measles and rubella risk and policy analysis. *Risk Analysis*, 36(7):1383–1403.

[Tompkins et al., 2009] Tompkins, D., Ramsey, D., Cross, M., Aldwell, F., De Lisle, G., and Buddle, B. (2009). Oral vaccination reduces the incidence of tuberculosis in free-living brushtail possums. *Proceedings of the Royal Society B: Biological Sciences*, 276(1669):2987–2995.

[Tompkins et al., 2013] Tompkins, D. M., Buddle, B. M., Whitford, J., Cross, M. L., Yates, G. F., Lambeth, M. R., and Nugent, G. (2013). Sustained protection against tuberculosis conferred to a wildlife host by single dose oral vaccination. *Vaccine*, 31(6):893–899.

[Valéry et al., 2008] Valéry, L., Fritz, H., Lefeuvre, J.-C., and Simberloff, D. (2008). In search of a real definition of the biological invasion phenomenon itself. *Biological invasions*, 10(8):1345–1351.

[Van Mieghem, 2011] Van Mieghem, P. (2011). The n-intertwined sis epidemic network model. *Computing*, 93(2-4):147–169.

[Van Mieghem et al., 2008] Van Mieghem, P., Omic, J., and Kooij, R. (2008). Virus spread in networks. *IEEE/ACM Transactions On Networking*, 17(1):1–14.

[Veale et al., 2012] Veale, A. J., Hannaford, O. D., Russell, J. C., and Clout, M. N. (2012). Modelling the distribution of stoats on new zealand offshore islands. *New Zealand Journal of Ecology*, pages 38–47.

[Veitch and Clout, 2001] Veitch, C. and Clout, M. (2001). Human dimensions in the management of invasive species in new zealand. *The great reshuffling: Human dimensions of invasive alien species*, pages 63–71.

[Vilà et al., 2010] Vilà, M., Basnou, C., Pyšek, P., Josefsson, M., Genovesi, P., Gollasch, S., Nentwig, W., Olenin, S., Roques, A., Roy, D., et al. (2010). How well do we understand the impacts of alien species on ecosystem services? a pan-european, cross-taxa assessment. *Frontiers in Ecology and the Environment*, 8(3):135–144.

[Vitousek et al., 1996] Vitousek, P. M., D'Antonio, C. M., Loope, L. L., and Westbrooks, R. (1996). Biological invasions as global environmental change. *American Scientist*, 84(5):468–478.

[Wang, 2010] Wang, F. (2010). Application of the lambert w function to the sir epidemic model. *The College Mathematics Journal*, 41(2):156–159.

[Wang et al., 2011] Wang, J., Shi, J., and Wei, J. (2011). Predator–prey system with strong allee effect in prey. *Journal of Mathematical Biology*, 62(3):291–331.

[Wang and González, 2009] Wang, P. and González, M. C. (2009). Understanding spatial connectivity of individuals with non-uniform population density. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1901):3321–3329.

[Wang et al., 2003] Wang, Y., Chakrabarti, D., Wang, C., and Faloutsos, C. (2003). Epidemic spreading in real networks: An eigenvalue viewpoint. *Proceedings of the IEEE Symposium on Reliable Distributed Systems*, (22):25–34.

[Warburton et al., 2009] Warburton, B., Cowan, P., and Shepherd, J. (2009). How many possums are now in new zealand following control and how many would there be without it. *Landcare research contract report LC0910/060*.

[Warburton and Livingstone, 2015] Warburton, B. and Livingstone, P. (2015). Managing and eradicating wildlife tuberculosis in new zealand. *New Zealand Veterinary Journal*, 63(sup1):77–88.

[Weeden and Cornwell, 2020] Weeden, K. A. and Cornwell, B. (2020). The small-world network of college classes: implications for epidemic spread on a university campus. *Sociological science*, 7:222–241.

[Whitford et al., 2014] Whitford, J., Rouco, C., Tompkins, D., and Nugent, G. (2014). First direct estimate of the detection probability of bovine tuberculosis in possums by possum transmission. *European Journal of Wildlife Research*, 60(5):827–830.

[Wilson et al., 1998] Wilson, P. R., Karl, B. J., Toft, R. J., Beggs, J. R., and Taylor, R. H. (1998). The role of introduced predators and competitors in the decline of kaka (nestor meridionalis) populations in new zealand. *Biological conservation*, 83(2):175–185.

[Wilson-Aggarwal et al., 2019] Wilson-Aggarwal, J. K., Ozella, L., Tizzoni, M., Cattuto, C., Swan, G. J., Moundai, T., Silk, M. J., Zingeser, J. A., and McDonald, R. A. (2019). High-resolution contact networks of free-ranging domestic dogs canis familiaris and implications for transmission of infection. *PLoS neglected tropical diseases*, 13(7):e0007565.

[Yang et al., 2015] Yang, W., Zhang, W., Kargbo, D., Yang, R., Chen, Y., Chen, Z., Kamara, A., Kargbo, B., Kandula, S., Karspeck, A., et al. (2015). Transmission network of the 2014–2015 ebola epidemic in sierra leone. *Journal of The Royal Society Interface*, 12(112):20150536.

[Youssef and Scoglio, 2011] Youssef, M. and Scoglio, C. (2011). An individual-based approach to sir epidemics in contact networks. *Journal of theoretical biology*, 283(1):136–144.

[Zhao et al., 2017] Zhao, W., Buffo, A., Alopaeus, V., Han, B., and Louhi-Kultanen, M. (2017). Application of the compartmental model to the gas–liquid precipitation of co2-ca (oh) 2 aqueous system in a stirred tank. *AIChE Journal*, 63(1):378–386.